

Open Research Online

The Open University's repository of research publications and other research outputs

Designing Activities for Collaboration at Classroom Scale Using Shared Technology

Thesis

How to cite:

Kreitmayer, Stefan (2015). Designing Activities for Collaboration at Classroom Scale Using Shared Technology. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2015 Stefan Kreitmayer



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000ab56>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Designing Activities for Collaboration at Classroom Scale Using Shared Technology

Stefan Kreitmayer

A thesis submitted to The Open University
for the degree of Doctor of Philosophy in Computing
30th September 2014

Department of Computing

Faculty of Mathematics, Computing and Technology

The Open University Milton Keynes, UK

Supervisors: Dr Robin Laney, Dr Stephen Peake, Prof Yvonne Rogers



Declaration

I hereby declare that the research presented in this dissertation is my own.

Abstract

Although researchers, teachers and policy makers broadly agree on the benefits of collaborative learning, there appears to be less clarity regarding how effective collaboration can be realised at classroom scale.

Research in Computer-Supported Collaborative Learning (CSCL), Human-Computer Interaction (HCI), simulation-based learning and related fields has produced a considerable range of applications that aim to support collaboration in classrooms. Grounded in well-established theories of how humans learn, many such applications have shown promising results within the context of small research studies. However, most of those research-driven applications never matured beyond the prototype stage and few are available today as products that schools can easily use and adopt. Many systems lack flexibility or require too much time, hardware, technical skills or other resources to be effectively implemented. Furthermore, teachers can be overwhelmed by managing large groups of students engaged in complex, computer-supported tasks.

This thesis investigates how forms of whole-classroom activity can be supported by combining shareable technologies with simulation, team play and orchestration. New designs are explored to help large groups engage and discuss at multiple scales (from pairs and small groups to the entire classroom) in ways that effectively include each student and use

the teacher's limited resources efficiently. Moreover, this research aims to devise and validate a conceptual framework that can guide future design, orchestration and evaluation of such activities. Three in-situ studies were conducted to address these goals.

The first study involved the design of a climate change simulation to support a professional training course. Iterative design and video analysis resulted in the formulation of the Collaborative Learning Orchestration for Verbal Engagement and Reflection (CLOVER) framework. This framework comprises a suite of conceptual tools and recommendations that aim to help designers and teachers create, orchestrate and evaluate decision-based simulations for whole-classroom use.

Two follow-up studies were conducted to validate the usability and usefulness of CLOVER. One of them aimed to replicate the previous findings in a similar context and resulted in the design of a sustainable, whole-classroom simulation for students to discuss finance decisions. The other used CLOVER to expand an existing desktop application (a language comprehension task for children) to classroom scale.

In sum, the three studies provide substantial empirical evidence, suggesting that CLOVER-based applications can effectively reconcile learning needs (collaboration) and technological affordances (shareable devices) with the inherent benefits and constraints of teacher-driven, co-located environments. Furthermore, the findings contribute to a better understanding of what it means to design for sustainability in this context.

Preface

Some of the materials in this dissertation also appear in the following conference papers:

1. Kreitmayer, Stefan; Rogers, Yvonne; Laney, Robin and Peake, Stephen. UniPad: Orchestrating Collaborative Activities Through Shared Tablets and An Integrated Wall Display. In: *Proceedings of the 2013 ACM Conference on Ubiquitous Computing (UbiComp '13)*, ACM, 2013, Zurich, Switzerland.
2. Kreitmayer, Stefan; Laney, Robin; Peake, Stephen and Rogers, Yvonne. Sharing Bubbles: Reflections on Offline Multi-Surface Scenarios. In: *Proceedings of the 2013 ACM Conference on Ubiquitous Computing (UbiComp '13)*, Contribution to the SOFTec workshop, ACM, 2013, Zurich, Switzerland.
3. Kreitmayer, Stefan; Rogers, Yvonne; Laney, Robin and Peake, Stephen. From participatory to contributory simulations: changing the game in the classroom. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12)*, pp. 49-58. ACM, 2012, Austin, Texas, USA.
4. Kreitmayer, Stefan; Peake, Stephen; Laney, Robin and Rogers, Yvonne (2011). Designing a Large Multi-Player Simulation Game to Encourage Reflection and Critical Debate. In: *Proceedings of the 5th European Conference on Game Based Learning (ECGBL '11)*, 20-21 Oct 2011, Athens, Greece.

Acknowledgements

This PhD research would not have been possible without the help of other people to whom I am greatly indebted.

The three studies reported in this thesis were collaborative projects, including teachers and other researchers. The 4Decades climate simulation would not have been possible without Stephen Peake (The Open University) who deserves credit as initiator, teacher and designer of the mathematics behind the simulation. Jörn Ketelsen programmed parts of the server application and the tablet interfaces. Vaiva Kalnikaitė contributed field notes during the first in-situ evaluation. Warm thanks to Richard Hearne (video) and Melissa Sutcliffe (photos) for their excellent camera work that supported my analysis. Ideas and support came from The University of Cambridge Institute for Sustainability Leadership (CISL) who hosted 4Decades at some of their professional training courses.

The UniPad simulation was designed in collaboration with Mike Mompi, Sharan Jaswal and other experts at MyBnk who contributed material, time and inspiration.

Nicola Yuill (University of Sussex) directed the preliminary evaluation of the Comfy Birds application at the ChaTLab and in schools. She and her students have been a constant source of wisdom, comfort and challenging questions.

To all the heroic participants in my studies: You deserve a medal. And cake.

Eternal gratitude goes to all the good people inside and outside of academia who have supported my personal transformation from being an idiot to being an idiot who wanted a PhD. Thank you all for helping me stay on the difficult path at times when easier paths were tempting.

Thanks to supervisors, mentors, proof readers, reviewers, role models, random folks, former teachers, friends and fellow PhD students and the fine people at The Open University, UCLIC (University College London), London Knowledge Lab and ChaTLab (University of Sussex).

Thanks to my parents for nature, nurture and the whole mess in between.

Also, thanks to animals.

Contents

| | |
|---|----------|
| 1. Introduction | 3 |
| 1.1. Motivation | 4 |
| 1.2. Background | 4 |
| 1.2.1. Rethinking classroom configurations | 5 |
| 1.2.2. Can shared technologies extend collaboration to the whole classroom? | 6 |
| 1.2.3. Realistic demands on the teacher | 7 |
| 1.3. Research questions and aims | 8 |
| 1.4. Approach | 8 |
| 1.4.1. Dealing with constraints: Towards feasibility in the wild | 10 |
| 1.4.2. Understanding classroom needs by aiming for adoption | 10 |
| 1.4.3. Transfer from professional training to classrooms | 11 |
| 1.4.4. Designing for sustainability | 12 |
| 1.5. Thesis outcomes | 14 |
| 1.6. Definitions | 16 |
| 1.6.1. Classroom | 16 |
| 1.6.2. Students / Learners / Participants | 16 |
| 1.6.3. Orchestration | 17 |
| 1.6.4. Classroom activity | 17 |
| 1.6.5. Teaching | 17 |

| | |
|---|-----------|
| 1.6.6. Simulation | 18 |
| 1.6.7. Models, simulations and games | 18 |
| 1.7. Thesis Structure | 19 |
| 2. Literature Review | 21 |
| 2.1. Introduction | 21 |
| 2.1.1. Aims of this chapter | 22 |
| 2.1.2. Overall strategy | 22 |
| 2.1.3. Chapter structure | 24 |
| 2.2. Part one: Theories of learning | 26 |
| 2.2.1. Higher level learning: Bloom's taxonomy | 26 |
| 2.2.2. Constructivist theories | 30 |
| 2.3. Part two: Simulation-based learning | 35 |
| 2.3.1. Common assumptions | 35 |
| 2.3.2. Relevant approaches | 44 |
| 2.3.3. First outcome: A lean pedagogy for applying theory in practice | 54 |
| 2.4. Part three: Computer-Supported Collaborative Learning | 57 |
| 2.4.1. Trends in the field | 57 |
| 2.4.2. Collaborative whole-class pedagogies | 59 |
| 2.4.3. What obstacles impede whole-class collaborative activities? | 60 |
| 2.4.4. Classroom orchestration | 61 |
| 2.4.5. Second Outcome: Roles of the teacher and technology | 65 |
| 2.5. Part four: Human-Computer Interaction | 67 |
| 2.5.1. New technologies | 67 |
| 2.5.2. The rise of touch-screen technologies in classrooms | 68 |
| 2.5.3. Extending from small-group HCI to large groups | 69 |
| 2.5.4. Problems with isolated sub-groups | 70 |
| 2.5.5. Strategies for improving linearity and awareness | 71 |

| | |
|--|-----------|
| 2.5.6. Third outcome: Starting points for interaction design | 73 |
| 2.6. Summary and conclusion | 74 |
| 3. Methodology | 75 |
| 3.1. Design-based research approach | 75 |
| 3.1.1. From concrete design to abstract concepts and back | 76 |
| 3.1.2. Design layers: The “Triple-i” Hierarchy | 78 |
| 3.1.3. Teachers taking part in the design | 79 |
| 3.1.4. Iterative design and prototyping | 80 |
| 3.2. Analytical focus | 82 |
| 3.2.1. Engaging, enjoyable, usable and useful | 82 |
| 3.2.2. Relevant metrics | 83 |
| 3.2.3. Why not use pre and post tests? | 83 |
| 3.2.4. Why not use control groups? | 85 |
| 3.3. Data collection | 86 |
| 3.3.1. Types of data | 86 |
| 3.3.2. Amount of data | 88 |
| 3.3.3. Lab tests versus in-situ evaluation | 88 |
| 3.3.4. Choice of settings: From professionals to schools | 90 |
| 3.4. Qualitative analysis | 93 |
| 3.4.1. Overview | 93 |
| 3.4.2. Scope | 94 |
| 3.5. Summary | 96 |
| 4. Designing a participatory simulation on climate change | 97 |
| 4.1. Overview | 97 |
| 4.1.1. Research question and aims | 100 |
| 4.1.2. Chapter Structure | 100 |

| | |
|---|-----|
| 4.2. Study context and approach | 100 |
| 4.2.1. Executive training courses | 101 |
| 4.2.2. Addressing a concrete learning need | 101 |
| 4.2.3. Research in practice: co-operation with the course committee . . . | 101 |
| 4.2.4. Whole-activity testing in the lab | 102 |
| 4.2.5. The teacher's role in the design | 103 |
| 4.2.6. Time Frame | 104 |
| 4.3. Study setting: The CLP course series | 104 |
| 4.3.1. Motivation for active learning | 104 |
| 4.3.2. Participants' educational backgrounds and needs | 106 |
| 4.3.3. Previous efforts towards applying theory in practice | 106 |
| 4.3.4. Ethnographical investigation | 107 |
| 4.4. Design method | 111 |
| 4.4.1. Gameplay Objectives | 111 |
| 4.4.2. Choice of topics | 112 |
| 4.4.3. Choice of evaluation strategies | 113 |
| 4.4.4. Devising an activity structure | 113 |
| 4.4.5. Choice of technology | 115 |
| 4.5. Introducing the three sketches | 115 |
| 4.6. First sketch: "Table of leaders" | 116 |
| 4.6.1. Design | 116 |
| 4.6.2. Evaluation | 118 |
| 4.6.3. Discussion | 120 |
| 4.6.4. Summary | 124 |
| 4.7. Second sketch: "Parallel futures" | 125 |
| 4.7.1. Design | 125 |
| 4.7.2. Evaluation | 129 |

| | |
|--|------------|
| 4.7.3. Discussion | 131 |
| 4.7.4. Summary | 132 |
| 4.8. Third sketch: “Four equal regions” | 133 |
| 4.8.1. Design | 133 |
| 4.8.2. Evaluation | 141 |
| 4.8.3. Discussion and changes | 147 |
| 4.8.4. Summary | 154 |
| 4.9. Overall discussion of the lab-based design | 154 |
| 4.9.1. Reflections on iterative design and evaluation | 154 |
| 4.9.2. Ecology of devices | 157 |
| 4.9.3. Suggested use qualities for ambient displays | 159 |
| 4.9.4. Ready for the wild? | 159 |
| 4.10. Chapter summary and conclusion | 161 |
| 5. Analysis 1: In-situ feasibility of the 4Decades simulation | 163 |
| 5.1. Overview of Chapters 5-7 | 163 |
| 5.1.1. Three analyses, four sessions, one study | 165 |
| 5.1.2. Research aims and method | 166 |
| 5.1.3. Setting | 166 |
| 5.1.4. Technical implementation | 167 |
| 5.2. Analysis 1: Overall feasibility | 169 |
| 5.2.1. Method | 170 |
| 5.2.2. Session structure | 170 |
| 5.2.3. Introduction and practice match | 173 |
| 5.2.4. Findings | 178 |
| 5.2.5. Discussion | 193 |
| 5.3. Chapter summary | 199 |

| | |
|--|------------|
| 6. Analysis 2: Changing simulation rules | 201 |
| 6.1. Question and aims | 202 |
| 6.2. Benefits of Rule-Changing | 202 |
| 6.3. Design iteration | 203 |
| 6.4. Design objectives: adapted from Gameplay | 204 |
| 6.4.1. Providing relevant choices of scenarios and winning conditions . . | 204 |
| 6.4.2. Devising adequate representations | 206 |
| 6.4.3. Giving participants a sense of the model author's perspective . . . | 206 |
| 6.4.4. Reusing tablets and ambient displays | 206 |
| 6.5. Changes to the application | 207 |
| 6.5.1. Tablet interfaces for rule changing | 207 |
| 6.5.2. An additional ambient display for the rules | 209 |
| 6.5.3. Changes to the team display | 211 |
| 6.6. Integrating rule-changing into the whole activity | 212 |
| 6.7. Lab testing | 212 |
| 6.8. Rejected ideas for the tablet interface | 212 |
| 6.9. Evaluation in the wild | 214 |
| 6.10. Method | 214 |
| 6.11. Setting | 215 |
| 6.12. Room layout and participant roles | 215 |
| 6.13. Procedure | 216 |
| 6.14. Findings | 216 |
| 6.14.1. Chosen rules | 217 |
| 6.14.2. Participants' reflections on the chosen rules | 218 |
| 6.14.3. Large groups sharing one tablet | 219 |
| 6.14.4. Dialogue during rule-changing | 220 |
| 6.14.5. Clarification among peers | 221 |

| | |
|---|------------|
| 6.14.6. Participants' overall ratings of the activity | 222 |
| 6.14.7. Comments in the debriefing | 222 |
| 6.14.8. The teacher's actions | 223 |
| 6.14.9. The teacher's reflections on rule-changing | 223 |
| 6.15. Discussion | 224 |
| 6.16. Reflections on the design objectives | 224 |
| 6.17. Reflections on the Benefits of Rule-Changing | 226 |
| 6.18. Usability | 228 |
| 6.19. Usefulness | 230 |
| 6.20. Desirability | 231 |
| 6.21. Conclusions from Analysis 2 | 232 |
| 7. Analysis 3: Continuous and equitable accessibility | 235 |
| 7.1. Overview | 235 |
| 7.2. Question and aims | 236 |
| 7.3. Method | 236 |
| 7.4. Findings | 237 |
| 7.5. Overview | 237 |
| 7.6. Session 1 | 239 |
| 7.6.1. Designed Layout | 239 |
| 7.6.2. Appropriation | 239 |
| 7.6.3. Issues | 239 |
| 7.7. Session 2 | 240 |
| 7.7.1. Designed layout | 241 |
| 7.7.2. Appropriation | 241 |
| 7.7.3. Issues | 241 |
| 7.8. Session 3 | 243 |
| 7.8.1. Designed Layout | 243 |

| | |
|--|-----|
| 7.8.2. Both teams: Problematic Corner-L formations | 243 |
| 7.8.3. Transformation in team A | 246 |
| 7.8.4. Implications for team A | 247 |
| 7.8.5. Transformation in team B | 248 |
| 7.8.6. Implications for team B | 249 |
| 7.8.7. Impact on participant experience | 250 |
| 7.8.8. Summary of session 3 | 252 |
| 7.9. Session 4 | 252 |
| 7.9.1. Designed Layout | 252 |
| 7.9.2. Appropriation | 253 |
| 7.9.3. High overall satisfaction | 253 |
| 7.10. Full overview of appropriation | 255 |
| 7.11. Discussion | 257 |
| 7.12. Effects of group formations on mechanisms of collaboration | 257 |
| 7.12.1. Implications of small-group formations | 257 |
| 7.12.2. Implications of large-group formations | 257 |
| 7.13. Recommendations for design and orchestration | 258 |
| 7.14. Explaining severe issues that caused exclusion | 260 |
| 7.15. Learning about inclusion by studying exclusion | 261 |
| 7.16. The Three Bubbles model | 263 |
| 7.16.1. Relation to proxemics | 265 |
| 7.16.2. Use in evaluation | 265 |
| 7.16.3. Use as a design template | 266 |
| 7.16.4. Use as a metaphor for shared user experience | 267 |
| 7.17. Summary of the third analysis | 268 |
| 7.18. Overall discussion of the three analyses | 268 |
| 7.19. Summary of Chapters 5-7 | 271 |

| | |
|--|----------------|
| 8. The CLOVER framework | 273 |
| 8.1. Overview | 273 |
| 8.1.1. Research questions | 273 |
| 8.1.2. Research aims | 273 |
| 8.1.3. Chapter structure | 275 |
| 8.1.4. Why call it a framework? | 275 |
| 8.2. The CLOVER Suite | 275 |
| 8.2.1. Tools | 277 |
| 8.2.2. Recommendations | 288 |
| 8.3. Conclusion | 292 |
| 9. Validating the framework in the wild | 293 |
| 9.1. Overview | 293 |
| 9.2. Questions and aims | 295 |
| 9.3. Chapter structure | 296 |
| 9.4. Method | 296 |
| 9.4.1. Assessing feasibility | 296 |
| 9.4.2. Assessing sustainability | 297 |
| 9.4.3. Assessing the use of CLOVER | 297 |
| 9.5. Overview of validation studies | 298 |
| 9.6. The UniPad study | 302 |
| 9.6.1. Background | 303 |
| 9.6.2. Aim of this study | 303 |
| 9.6.3. Setting | 303 |
| 9.6.4. Approach to design and evaluation | 303 |
| 9.6.5. The application | 304 |
| 9.6.6. CLOVER-based design process | 309 |
| 9.6.7. In-situ feasibility | 313 |

| | |
|---|------------|
| 9.6.8. Discussion | 320 |
| 9.6.9. Conclusions from the UniPad study | 328 |
| 9.7. The ComfyBirds study | 330 |
| 9.7.1. Aim of this study | 330 |
| 9.7.2. Method and setting | 330 |
| 9.7.3. Overview of the design and analysis | 331 |
| 9.7.4. The learning task: Joking riddles | 331 |
| 9.7.5. CLOVER-based design rationale and objectives | 334 |
| 9.7.6. Classroom prototype | 338 |
| 9.7.7. Evaluation | 343 |
| 9.7.8. Potential for design iterations | 347 |
| 9.7.9. Overall sustainability | 348 |
| 9.7.10. Reflecting on CLOVER | 349 |
| 9.7.11. Conclusions from the ComfyBirds study | 349 |
| 9.8. Discussion | 350 |
| 9.8.1. Multi-Plane Groupwork tends to surprise | 350 |
| 9.8.2. Varieties of Three Bubbles | 351 |
| 9.8.3. Potential for extending CLOVER | 351 |
| 9.9. Conclusion | 353 |
| A. 4Decades rule-changing discussions | 355 |
| B. How to use CLOVER | 361 |
| B.1. Letting CLOVER guide a new design | 361 |
| B.2. Other ways to use CLOVER | 368 |
| Bibliography | 369 |

Chapter 1.

Introduction

This thesis investigates new ways of supporting collaborative learning at classroom scale through the use of shared technology. It integrates perspectives from multiple areas of literature; most importantly collaborative learning, games and simulations, and human-computer interaction (HCI). The main contribution is a conceptual framework titled Collaborative Learning Orchestration for Verbal Engagement and Reflection (CLOVER).

CLOVER aims to provide conceptual tools and recommendations that can guide the design and evaluation of whole-class, collaborative activities. Although CLOVER's primary focus is on simulations, some parts of it are generic to a wider range of activity structures. Indeed, out of the three empirical studies described here that underpin the framework, two are simulations and one is not. Since CLOVER is essentially a *set* of independent components, rather than a monolithic framework, it is possible to apply parts of CLOVER selectively, using only the elements that are relevant to a particular case.

Moreover, some of CLOVER's elements (namely the conceptual tools) come in a particular order, starting with fundamental definitions and moving towards more specific and optional features. When used to guide a new design from scratch, this sequence is intended to help designers ask the right questions at the right time. The example in Appendix B demonstrates how this can be used in practice.

This first chapter introduces the motivation, background, research questions, aims and overall approach. Key terms are defined and the structure of this thesis is outlined at the end of the chapter.

1.1. Motivation

It is widely accepted that collaborating with peers offers unique benefits for learning. Relevant social interactions - such as explaining, disagreement and mutual regulation - can trigger additional cognitive mechanisms, including knowledge elicitation, internalisation, and reduced cognitive load (Dillenbourg 1999, p.5).

These benefits have been shown repeatedly in lab-based studies with pairs or triplets using computer simulations together and discussing (e.g. Stahl 2006). However, what works well in small and isolated settings does not always transfer easily to real situations with larger groups, such as typical classrooms and professional training. Recent work has found that some, but not all teachers are confident in designing and orchestrating collaborative activities with large groups (De Hei et al. 2014; Moher et al. 2005).

While some digital tools can facilitate collaboration in principle, they often demand too many resources to be used broadly under real conditions (Sharples 2013). Besides the obvious hardware and software requirements (which can be prohibitive per se) studies have emphasised that teachers have limited time for preparation and maintenance (Faria and Wellington 2004) and a limited capacity for dividing their attention between collaborating teams (Alavi et al. 2009; Kharrufa et al. 2013).

Recent progress has been made in understanding what these constraints and their implications are (Dillenbourg and Jermann 2010). Nevertheless, transforming this knowledge into practical approaches remains a big challenge for research and design. The approach described in this thesis focuses on whole-class simulation and discussion. It aims to be practically feasible and sustainable in real settings.

1.2. Background

Recent research in collaborative learning has aimed to improve learning through active engagement, participation in groups, frequent interaction and feedback, and connections to real-world contexts (Roschelle et al. 2000). Toward this goal, established theories of collaborative learning have been taken out of the lab and applied in the design of new tools that are evaluated in real-life classrooms. Using this approach of educational research in the wild, a considerable body of tools and empirical evidence has been produced (e.g. Dillenbourg et al. 2012).

1.2.1. Rethinking classroom configurations

Tom Moher's RoomQuake simulation for secondary science classrooms (Moher et al. 2005) demonstrates how collaborative learning can be conducted in classrooms by carefully reconfiguring the roles of students, teachers and technology (Figure 1.1). Using a mix of digital and non-digital materials and authentic scientific methods, students measure and analyse seismic events that are simulated within the space of the classroom. Similar simulations have been designed around various science topics under the umbrella term of Embedded Phenomena (Moher 2006).



Figure 1.1.: Children working together in a science class, using the RoomQuake application.

By encouraging the whole class to walk around in the whole room, talk and share devices in groups, and regroup opportunistically, Moher's work proposes a different approach of using technology in schools. It is in many ways contrary to the typical school's computer lab that offers little support for face-to-face collaboration, as it is configured more or less like a call centre, where students are largely isolated (Fig. 1.2).



Figure 1.2.: Laptops in a computer lab

1.2.2. Can shared technologies extend collaboration to the whole classroom?

Alternative classroom set-ups have been proposed, such as the SynergyNet classroom (AlAgha et al. 2010) that capitalises on shareable technologies to encourage collaboration among students. As Figure 1.3 shows, SynergyNet features a network of multi-touch tables and interactive whiteboards to help students work together in groups, whilst allowing the teacher to interact easily with the groups. Large displays provide integrated information to the teacher and the whole class. Prototypical evaluations within controlled settings have shown that such multi-surface systems can support small-group collaboration, flexible sharing between groups and the teacher's ability to orchestrate the interaction (Higgins et al. 2011).



Figure 1.3.: SynergyNet: (a) Up to four students can collaborate at each table. (b) The teacher can monitor and control the students' tables. (c) The table contents can be shared on a vertical screen for discussion among the whole class.

Such promising findings give rise to the question whether systems like SynergyNet will be feasible in the wild. So far they have been conducted largely in lab-based studies, providing new insights into aspects of collaborative learning. However, are such systems practically ready for large-scale adoption in schools? Hardware cost and maintainability are obvious concerns in this respect. Aggravating factors include potential problems of legibility and coordination at each tabletop (Morris 2005), especially when the horizontal screens are approached from different angles (Wigdor and Balakrishnan 2005). These issues are typical for tabletop systems. Further obstacles exist that are generic to classroom tools and pedagogies.

1.2.3. Realistic demands on the teacher

When teachers ponder the adoption of new technologies in classroom, they are confronted with realistic questions about resources feasibility. For example, would the planned activity require the teacher to:

- set up a server, network or other advanced computing infrastructure?
- use any non-standard hardware or software?
- install new software on school computers?
- learn new ICT skills?
- obtain or produce any physical or digital materials?
- assist students with using the interface of an application?
- prepare the room for the activity?
- etc.

If the answer to any of those questions is yes - as is the case with many prototypical designs for whole-classroom collaboration, including participatory simulations (Wilensky and Stroup 1999; Colella 2000) and Embedded Phenomena (Moher 2006) - then there may be significant implications for time demand, logistical effort, getting permissions, etc. The sum of such barriers can make an activity less feasible in the wild and potentially hamper its progression from prototype stage into large-scale use.

1.3. Research questions and aims

The goal of this thesis is to determine how practically feasible forms of whole-classroom activity can be supported by combining shareable technologies with simulation, team play and orchestration. To address this, the following research questions are asked:

- **Q1.** How can whole-classroom activities be designed that use shared technology to facilitate large groups' problem solving and debate in a face-to-face setting?
- **Q2.** What factors are instrumental in orchestrating such activities in the wild?
- **Q3.** What abstractions can we identify that would form a helpful conceptual framework for activity design and orchestration?
- **Q4.** Can this framework generally support the design of sustainable classroom activities?

The two overarching aims of this dissertation are:

1. to investigate new ways of promoting large-groups' collaborative learning, by designing and evaluating classroom-scale activities. The focus is on enabling, encouraging and facilitating joint exploration of problem spaces, debate and reflection.
2. to devise and validate a generalisable framework that can guide future design, orchestration and evaluation of whole-class collaborative activities.

1.4. Approach

The research questions were addressed through an iterative approach, including the development and validation of a conceptual framework (Fig. 1.4). The first study addressed Q1 by exploring how shareable technologies can support collaborative learning goals in a whole-class simulation (blue). Professional training workshops were chosen as a setting for this initial study, to allow multiple consecutive deployments in real situations. This is where Q2 was addressed. The findings led to the formulation of a conceptual framework to guide the design, orchestration and evaluation of similar whole-class activities, in response to Q3. This framework was tested and refined in two further studies (green) to evaluate whether teachers and researchers find the framework useful, and to test how sustainable the approach is in restricted settings. These final studies addressed Q4 and provided additional support for Q1, Q2 and Q3.

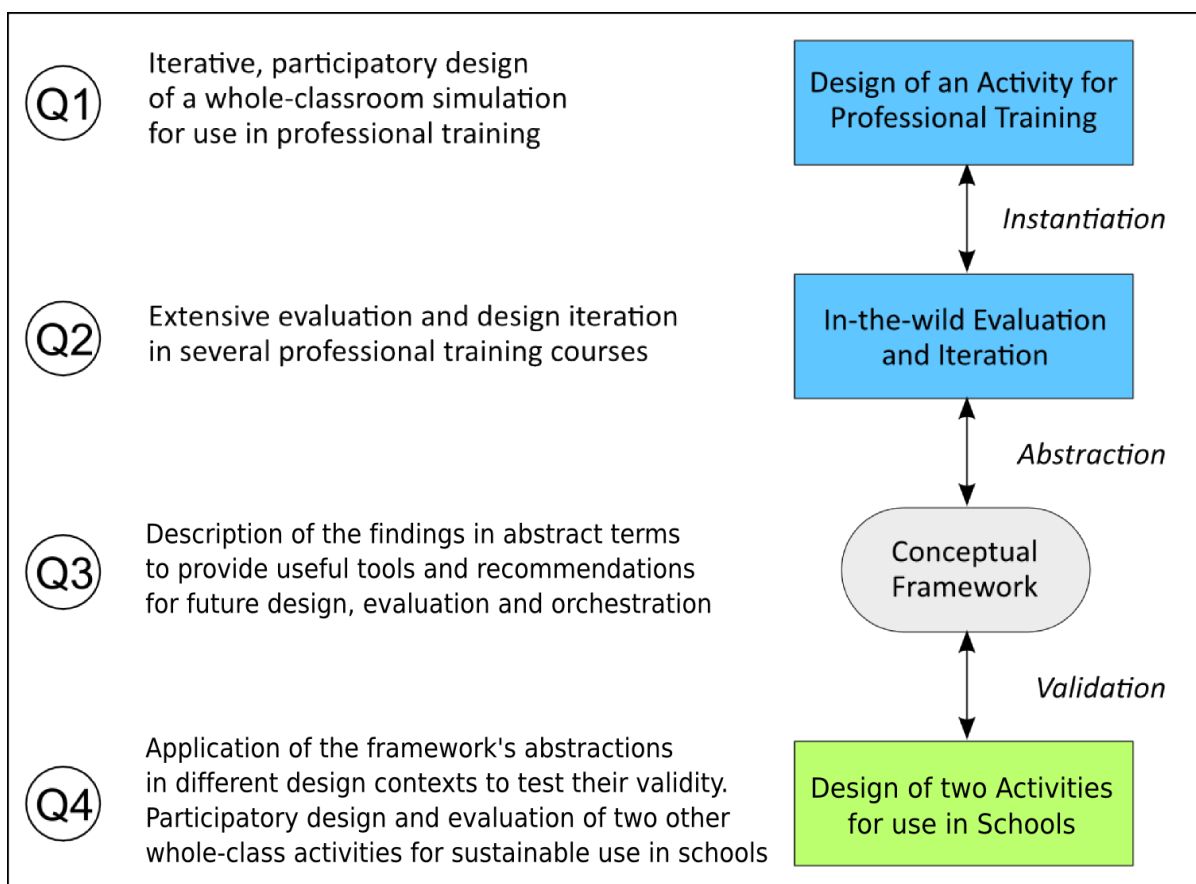


Figure 1.4.: Overview of the approach taken in this research

1.4.1. Dealing with constraints: Towards feasibility in the wild

In preparation to the main studies, a literature was conducted to identify (a) what factors have so far hampered advancement of collaborative pedagogies in classroom settings, and (b) where potential exists for technology to improve the situation.

A variety of relevant obstacles were revealed through the literature review, including practical, theoretical, cultural, political, economic and organisational factors. Classroom-typical constraints include time, assessment, curriculum, energy, space and safety (Dillenbourg and Jermann 2010). Collaborative tools and activities, in order to be practically feasibility, must be flexible enough for the average teacher to use within these constraints in real time Dillenbourg et al. (2012).

The review further revealed potential for design in the areas of shared interfaces, team-based role play, and an integrated pedagogy involving traditional lecture, simulation and discussion.

1.4.2. Understanding classroom needs by aiming for adoption

This PhD project identifies with a contemporary movement in Technology-Enhanced Learning (TEL) that has aimed to create tools that are feasible in the wild, i.e., outside experimental settings:

“Many innovators [...] want to move beyond doing good science that merely promises later application; [they] now find educational needs (and not just related scientific problems) to be truly motivating and want to create materials and tools with impacts that can be realized in their lifetimes in practical educational settings at scale.” (Roschelle et al. 2013)

“[...] reflects a rethinking of the relationship between research and practice, seeking ways to design learning tools so that they are theoretically motivated, practically usable and useful, and designed for the realistic classroom contingencies that require flexible adaptations by teachers.” (Dillenbourg et al. 2012)

For instance, Tom Moher’s research group, having published extensively in learning sciences as well as Human-Computer Interaction (Moher 2006; Moher et al. 2010), has gone to considerable trouble to make their RoomQuake classroom simulation accessible to teachers everywhere. This included the software as well as detailed instructions for teachers that were carefully prepared and offered for download: <http://dkilb8.wix.com/roomquake>

(Accessed on 14 April 2011.) These materials demonstrate a sincere commitment to engaging with teachers' needs on all levels, taking the very mundane technical and logistical details as seriously as the high-level pedagogical intentions. Several PhD students in the group are currently dedicated to making Embedded Phenomena more accessible to the public.

1.4.3. Transfer from professional training to classrooms

The conceptual framework was initially developed in professional training settings and later transferred to school settings where it was refined. The motivation for doing it in this order, rather than the other way round, is detailed in the methodology (Chapter 3). It can be summarised as follows:

Firstly, professional training settings are supported by a wealth of previous literature in the area of simulation-based and team-based learning. Being able to build on well-established assumptions and “good practices” in this area promised to be a good starting point.

Secondly, the chosen professional training setting enabled a more “agile” and exploratory approach to activity design than is normally possible in a school setting. Key benefits included plenty of preparation time, multiple evaluations per year and a fundamental, organisation-wide commitment to collaborative learning. These benefits would allow the design to focus on participants' essential learning experience, rather than peripheral restrictions.

Thirdly, demographics played an important role in the choice of settings. Participants in the first study were mature and likely to be already skilled in collaborative working. To gradually explore the boundaries of the CLOVER approach with regard to demographics, the second study involved pre-university students and the third study involved children with learning difficulties.

To summarise, the main reasons concerned the available literature as well as methodological considerations. In both respects, the transition to school settings implied an increase of difficulty, due to (a) less previous work to build on and (b) a stricter set of practical constraints. Hence, the second and third study can be seen as more ambitious than the first study - for the above reasons and also because the latter studies aimed to create, not just a *feasible* but a *sustainable* approach (see Figure 1.5 further below).

1.4.4. Designing for sustainability

The research strategy was to evaluate the designed activities by conducting them *repeatedly* in real settings. In contrast to the common practice of one-off evaluation (e.g. Moher et al. 2005), this strategy goes beyond testing the usability, usefulness and desirability of a pedagogical approach; it furthermore aims to test its *sustainability*. To be considered sustainable, an approach needs to have the potential for sustained use under everyday conditions, i.e., requiring only the amount of commitment that a teacher is ready to summon up on a regular basis. This amount is assumed to be less than the special efforts that teachers may be willing to make at the special occasion of a scientific experiment involving some exciting new technology.

The goal of sustainability imposes an additional layer of rigour on the test conditions. Particularly, the observation that an approach “worked well” once under the guidance of researchers is not enough. Rather, there should be evidence of teachers (or institutions) being able to apply the approach repeatedly on their own, under everyday conditions.

The total effort a teacher may invest (preparation work, time to provide feedback, etc) is limited. If we design for heroes, we lose scalability (there are few heroes) and sustainability (heroes get tired). (Dillenbourg et al. 2012)

A distinction is made between sustainability and *adoption* (i.e. actual sustained use), since the latter may depend on further constraints, such as institutional strategy, politics and management. Figure 1.5 illustrates these concepts as a hierarchy, whereby the concepts near the bottom represent requirements for those further above.

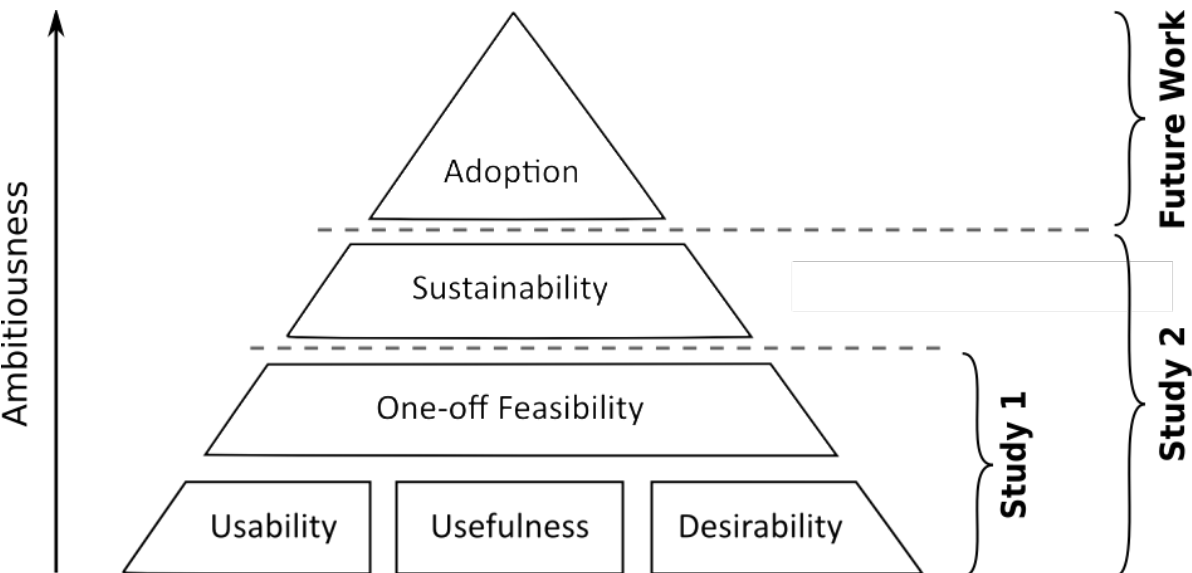


Figure 1.5.: Figure : Design goals of this thesis included sustainability, but not adoption.

1.5. Thesis outcomes

This dissertation proposes a new approach to engaging a whole class in joint problem solving and debate. It describes how to design whole-class, face-to-face activities that any teacher can set up within a few minutes and facilitate independently of external help (particularly that of researchers). This is achieved by the following main outcomes:

1. Three case studies demonstrating the feasibility and sustainability of the approach;
2. The Collaborative Learning Orchestration for Verbal Engagement and Reflection (CLOVER) conceptual framework that was developed and validated as a result of these studies.

While the case studies contribute empirical knowledge regarding what works and what doesn't in the wild, the framework comprises a set of abstractions (conceptual tools and recommendations) to make this knowledge accessible in new design situations. Particularly, CLOVER aims to support the design and evaluation of large-group, collaborative activities for use in any "classroom-like" setting, including school classrooms, university seminars and professional training workshops.

The proposed approach aims to be generally applicable to a broad range of pedagogical contexts, topics and learner populations. The three empirical studies illustrate this range.

1. The 4Decades simulation of future climate change that adult participants can critique and modify in situ;
2. The UniPad simulation for pre-university students that any teacher can rapidly set up and orchestrate within the last 15-25 minutes of a lesson;
3. The ComfyBirds whole-classroom activity to support reading comprehension by discussing ambiguous words in joking riddles.

All three applications were shown to encourage remarkable instances of collaborative learning - including joint problem solving, critical debate and peer teaching - thus enabling participants to instantly apply theory in practice. This is achieved through six core design principles that emerged from the iterative process.

- Practical choices of available technology: shared tablets and projectors;
- Simple tablet interfaces that do not require assistance or time to learn;
- An emphasis on verbal participation, rather than complex multi-touch interaction;

- Fluid constraints on group interaction via familiar social mechanisms, such as sharing, team play and social comparison;
- Appropriate interaction techniques to limit the teacher's workload during the activity;
- Browser-based client software that is platform independent and easy to access - no native software installation or special hardware required.

The success of these applications and the shown generalisability of CLOVER indicate wider potential for future design in many areas of formal education and informal learning.

1.6. Definitions

This dissertation relies on a number of key concepts which are defined below.

1.6.1. Classroom

The word “classroom” is used here in an overarching sense, referring to any co-located situation where “*a person (teacher, teaching assistant, parent, workplace supervisor, etc.) has the responsibility to bring other persons to reach learning goals*” (Dillenbourg et al. 2011). This definition includes the typical school classroom as well as many professional training settings and university seminars. Furthermore, the studies described here emphasise **large-group** settings, with typically 20 to 30 learners per teacher, as is common for schools in the UK and North America (Blatchford et al. 2011).

1.6.2. Students / Learners / Participants

The words “student” and “participant” are used interchangeably in this dissertation. Strictly speaking, it would be most accurate to use “participant” throughout, since the focus of analysis in all studies was people’s *participation* in learning activities. The pedagogical approach described here does not distinguish as to whether participants have formal student status, whether they are in higher education, employment, and so on. Indeed, it is based on studies that have considered pre-university students (Kreitmayer et al. 2013b) as well as adults in professional and postgraduate training (Kreitmayer et al. 2012) and children in special education (Yuill et al. 2014). Sometimes I say “student” rather than “participant” simply to avoid repetition or to emphasise aspects that concern an individual (or group) at a wider scope, beyond a specific learning activity.

Similarly, the word “learner” appears occasionally when “student” would sound too specific. This is the case, e.g., when discussing learning theories that are framed more universally and not limited to the context of formal education (e.g. Vygotsky 1978; Lave and Wenger 1991; Stahl 2006). This wider notion of *people engaged in learning* is useful sometimes, since it includes students, professionals and others. However, on the flipside there is a high risk of ambiguity, since the “learner” label theoretically applies to any living person. To avoid confusion between participants’ learning and what researchers, designers and teachers learned in the context of the studies, this dissertation mostly uses the word “learner” sparingly and only when the meaning is unambiguous.

1.6.3. Orchestration

The term “orchestration” has emerged in the wider context of classroom research as a metaphor for good instructional design and lesson enactment in real classrooms (Looi and Toh 2014). Research under this label is characterised by an analytical focus on the classroom as a whole - rather than, e.g., mental processes of the individual learner, and being amenable to the complexities entailed by collaboration and interactive technologies. Dillenbourg et al. (2011) speak of classroom orchestration as *“a question of usability in which the classroom is the user”*.

1.6.4. Classroom activity

The research focus of this thesis is on face-to-face activities that aim to engage a whole class simultaneously. This excludes activities that are predominantly remote, screen based and/or asynchronous, such as homework, virtual learning environments (VLEs), single-user games and virtual worlds.

1.6.5. Teaching

Teaching is conceptualised here essentially as learning facilitation within a co-located lesson, by integrating collaborative activities effectively with lectures. Two fundamental assumptions underlie all of the studies described here:

1. People tend to learn better when given the opportunity to apply their knowledge in practice, for instance through argumentation (Jonassen and Kim 2010) or problem solving (Gardner 1991; Gee 2003).
2. People can practice more effectively if some instrumental background knowledge has been established previously, e.g., through introductory lectures (Silvia 2012).

Engaging collaboratively with specific problems through discussion and shared problem solving was defined as the primary learning objective for participants. Helping them achieve this was seen as the overall goal of design and orchestration.

1.6.6. Simulation

A simulation is any attempt to mimic a real or imaginary environment or system (Rieber 1996).

In adult education, role-play and simulations are commonly used across fields as diverse as medicine (Fanning and Gaba 2007), business (Faria et al. 2009) international relations (Morgan 2003), aviation, military, crisis management and many others. Simulations can be useful teaching tools because they allow real problem solving and collaboration in safe (simulated) environments. For this purpose, numerous simulations are available as products that can be bought or hired. Some offers include specially trained facilitators travelling to the training site.

By contrast, in school classrooms simulations are the exception. To explore possible reasons and remedies, this dissertation reviews a variety of approaches, learning goals and practical challenges in both school and adult education.

1.6.7. Models, simulations and games

Eykhoff (1974) defines a model as *“a representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in usable form”*. By emphasising “the essential aspects”, this definition implies that a model is always a reduced version of the real thing.

Whereas a **model** can be represented in a static form, e.g., mathematical equations or source code, a **simulation** can be thought of as the execution of a model over time. The execution may be interactive or non-interactive (such as, e.g., an animated weather map). If a simulation involves human interaction *and a winning condition* then it can also be referred to as a **game** by common definitions (Juul 2003) - regardless of the notion that some games are not simulations. These concepts are illustrated in Figure 1.6.

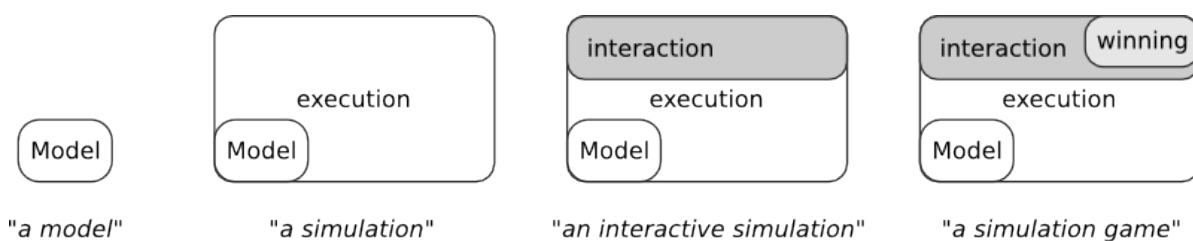


Figure 1.6.: Distinction between models, simulations and games.

1.7. Thesis Structure

Chapter 2 is the literature review and Chapter 3 explains the methodology. Chapter 4 addresses the first research question based on iterative prototyping in the lab. Chapters 5-7 cover the in-situ evaluations of 4Decades. The resulting framework is presented and validated in Chapters 8 and 9. The last two chapters are discussion and conclusion.

Chapter 4: Designing a participatory simulation on climate change. This chapter details the iterative design of the 4Decades simulation for use in professional training courses. The design process is described, involving several sketches, prototypes and lab tests.

Chapters 5-7: In-situ evaluation of the climate simulation. Three chapters are dedicated to the in-situ evaluation of the 4Decades simulation in professional training workshops. Three stages of qualitative analysis were conducted in order to (1) verify the overall feasibility of the approach, (2) explore potential for groups to make informed changes to the simulation and (3) examine the impact of spatial configurations (of peers and technology) on equitable participation among the group.

Chapter 8: Developing a conceptual framework. This chapter introduces the Collaborative Learning Orchestration for Verbal Engagement and Reflection (CLOVER) framework. It emerged out of the 4Decades empirical studies and aims to provide a suite of useful abstractions (conceptual tools and recommendations) to inform future design, orchestration and evaluation.

Chapter 9: Validating the framework in the wild. This chapter describes two more studies that provide empirical support for the usability and usefulness of CLOVER in different design situations where the aim was sustainable use in classrooms. The first study was a new design from scratch - the UniPad finance simulation. The second study involved a different design challenge whereby an existing desktop application was transformed into the ComfyBirds whole-class activity.

Chapter 10: Discussion. This chapter discusses the overall research findings within the context of the literature.

Chapter 11: Conclusions and future work. This chapter summarises the research contributions, considers limitations of this research, discusses work in progress and suggests perspectives for future research.

Chapter 2.

Literature Review

2.1. Introduction

The work described in this dissertation is grounded in multiple threads of related research. Besides foundational theories of learning, the main influences came from the following three areas of applied research:

- research on learning with simulations and games, including computer simulations
- the field of Computer-Supported Collaborative Learning (CSCL)
- the field of Human-Computer Interaction (HCI)

Each of these fields bears a range of theories, research methods and empirical work relevant to the design, orchestration and evaluation of classroom-scale, face-to-face collaboration. Relevant insights from each of these fields are discussed in this chapter.

2.1.1. Aims of this chapter

The aims of this chapter are:

- to survey common assumptions regarding how simulations and collaboration can support learning and teaching;
- to identify which of these assumptions are potentially relevant to classroom contexts;
- to devise a well-informed pedagogical approach for classrooms based on these assumptions;
- to review previous classroom-based approaches that involve simulation and collaboration and evaluate those approaches in regard to their feasibility and sustainability in typical, classroom-type settings;
- to highlight research and design challenges, practical obstacles, pitfalls, good practices and promising strategies;
- to devise useful starting points for instructional design, interaction design, interface design and evaluation of simulation-based, whole-classroom collaboration.

2.1.2. Overall strategy

To approach these aims, four main fields of literature were investigated: (1) general theories of human learning, (2) learning with simulations, including computer simulations and games, (3) computer-supported collaborative learning and (4) human-computer interaction.

Figure 2.1 illustrates these four fields and foreshadows the sequence in which they will be introduced in this chapter. This sequence is supported by a series of key questions that are associated with these areas. The logical order of these questions describes an overall transition from abstract (theoretical) to concrete (practical) considerations. Indeed, this is a deliberate strategy, guided by the overarching goal of this dissertation to devise practical solutions that are grounded in theory. Starting with theoretical foundations of learning, the discussion will continue to address practical aspects of teaching (involving simulations and collaboration) and eventually go into detail regarding design and technology.

| | |
|---|--|
| PART 1 <hr/> Theories of Learning <hr/> Section 2.2 | What theories of human learning are most relevant to the aims of this dissertation? Can we reconcile perspectives on direct instruction, collaboration and simulation? How can discussion and argumentation be leveraged as pedagogical tools? |
| PART 2 <hr/> Simulation- based learning <hr/> Section 2.3 | What are the relevant benefits of simulations for learning and teaching? What are realistic learning objectives in a classroom context? What are the instructional requirements, good practices, pitfalls, etc.? |
| PART 3 <hr/> CSCL / Orchestration <hr/> Section 2.4 | Which learning goals are feasible within the constraints of a classroom setting? What are the realistic capacities of one teacher within a large-group simulation? How can technology and a human teacher complement each other efficiently? |
| PART 4 <hr/> HCI <hr/> Section 2.5 | How can current designs be improved in terms of participation and orchestration? What technologies are most appropriate? What are useful starting points for interaction design and interface design? |

Figure 2.1.: Overview of the four parts of the literature review (Sections 2.2 - 2.5 in this chapter)

2.1.3. Chapter structure

Following the above logic, the remainder of this chapter is sectioned into four main parts, each focussing on one of the research areas:

- Part 1: Learning theories (Section 2.2)
- Part 2: Simulation-based learning (Section 2.3)
- Part 3: Computer-supported collaborative learning (Section 2.4)
- Part 4: Human-computer interaction (Section 2.5)

Each of these sections finishes with a set of preliminary conclusions that are summarised in the final conclusion section at the end of this chapter.

Needless to say, the fields are densely interrelated, rather than cut and dried. Consequently, some of the key guiding questions relate to more than one field and it is only realistic that the review will make many cross-references.

Part 1: Theories of learning (Section 2.2) This section sets the scene by introducing some general theories of “active” learning, based on research in education and cognitive science. The main ideas are (a) Bloom’s distinction between “higher level” and “lower level” learning objectives and (b) a variety of constructivist learning models, including Kolb’s experiential learning cycle and collaborative learning theories.

Part 2: Simulation-based learning (Section 2.3) The second part focuses on the use of simulations to support learning goals, such as engagement, conceptual understanding, productive discourse, skill development and literacy. Common assumptions about these learning goals are discussed in relation to classroom settings, resulting in a tentative proposal for how to effectively integrate traditional lectures, simulation and discussion in the context of a lesson or seminar. The proposal is formulated as a generic model to help instructors link together the benefits of each component (e.g., the benefit of simulations to trigger productive discourse). It does, however, not provide any concrete explanations about how these benefits can be achieved, e.g., how to design a simulation that can engage a whole classroom, what kinds of technology are appropriate or how to go about facilitation and debriefing.

Part 3: Computer-supported collaborative learning (Section 2.4) In the third part of the review, the CSCL literature is consulted to explore more concretely how the proposed approach can be realised in actual classrooms. Focusing on aspects of orchestration, this part of the review investigates what the specific capacities of technology

and a single human teacher are in managing large-group collaborative learning under real-life classroom conditions. Selected approaches are discussed critically, leading to a number of tentative conclusions regarding how design can effectively reduce the teacher's orchestration load, rather than making things more complicated. The result is a clearer picture of what the roles of the teacher and technology should be and how they can complement each other. However, questions remain as to which technologies to use and how to design activities and interfaces. These questions are addressed in the next section.

Part 4: Human-Computer Interaction (Section 2.5) The fourth and final part takes a step back from the education-centred tradition of CSCL and assumes a more generic, HCI-centred perspective, focusing on how groups of people can engage with peers and technology. Relevant studies are revisited with regard to choices of technology, social mechanisms and interaction types. The goal is to explore potential for more engaging, portable and scalable designs. One of the main conclusions from this section is that ambient displays and shareable touch-screens bear much potential for innovative interaction design, by allowing participants to walk around and interact opportunistically with the technology and peers. Tentative design recommendations are outlined regarding how to effectively constrain information, control and awareness to support fluid collaboration in the whole group.

Summary and conclusion (Section 2.6) summarises the conclusions from the previous sections, resulting in a set of working assumptions that provide the foundation for the design-based research studies reported in the subsequent chapters.

2.2. Part one: Theories of learning

There are many theories of learning. The ones discussed here have been selected due to their significance to this work. All of them emphasise “active” learning and some focus specifically on collaboration.

2.2.1. Higher level learning: Bloom’s taxonomy

Bloom’s well-known “taxonomy of educational objectives” (1956) proposes a hierarchy of six categories to describe a comprehensive spectrum of educational goals: (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5) Synthesis and (6) Evaluation. Bloom refers to the first two categories (1) and (2) as “lower level”, and (3-6) as “higher level” learning objectives.

This model has been widely cited and translated. According to Krathwohl (2002), it deserves credit for directing awareness to a wider range of learning objectives in a landscape of formal education that had previously been (and arguably still is) overly focused on the lower levels.

2.2.1.1. The revised version of Bloom’s taxonomy

The revised version of the taxonomy differs in significant ways. Most strikingly, the names of the six categories were changed from nouns to verbs (e.g., “knowledge” was changed to “remember”; “comprehension” was changed to “understand”, and so on). Moreover, the two highest categories were swapped and renamed, and a second dimension was added to differentiate between different types of knowledge: factual, conceptual, procedural and metacognitive (Figure 2.2).

Several aspects of the original taxonomy have been criticised and were changed as part of the revision. Most significantly, Bloom’s original claim that each category requires each lower category was somewhat relaxed in the revised version due to (a) lack of empirical support and (b) the concession that teaching practices may vary. To paraphrase Krathwohl (2002), while the six categories on the whole describe a spectrum from simpler to more complex cognitive processes, there is much overlap and ambiguity between some of the categories, especially neighbouring ones. Indeed, two neighbouring categories were swapped in the revised version, which further indicates that the hierarchical dependencies are really not as clear and strict as originally postulated.

| The Knowledge Dimension | The Cognitive Process Dimension | | | | | |
|-----------------------------------|---------------------------------|----------------------|-----------------|-------------------|--------------------|------------------|
| | 1. <i>Remember</i> | 2. <i>Understand</i> | 3. <i>Apply</i> | 4. <i>Analyze</i> | 5. <i>Evaluate</i> | 6. <i>Create</i> |
| A. <i>Factual Knowledge</i> | | | | | | |
| B. <i>Conceptual Knowledge</i> | | | | | | |
| C. <i>Procedural Knowledge</i> | | | | | | |
| D. <i>Metacognitive Knowledge</i> | | | | | | |

Figure 2.2.: The revised version of Bloom's taxonomy (Krathwohl 2002)

Despite these modifications, the general distinction between *lower level* and *higher level* cognitive processes was preserved in the revised version and has been widely accepted among scholars. This dichotomy has appeared in various flavours across the literature, such as *rote learning* versus *meaningful learning* (Mayer 2002) or *content learning* versus *higher-level learning* (Fellenz 2004). These are essentially synonyms or variants with slightly different connotations. Many authors agree that the higher level objectives are important goals of education (Krathwohl 2002). However, there is much debate regarding how to apply the taxonomy in pedagogical practice - for instance, whether instruction should first focus on the lower levels before addressing the higher levels, or vice versa (Wegerif 2006).

Although Bloom's taxonomy is not in itself specific to collaborative learning, some of its key ideas are still being discussed in collaborative learning research (e.g. Wegerif 2006; Echeverría et al. 2011). From within this discourse, this dissertation adopts two key assumptions:

1. It is useful to distinguish between high and low level learning.
2. Some teaching methods tend to support the high-level cognitive processes better than others. For instance, active learning techniques (such as simulations) are well suited to the higher levels, whereas lectures and reading assignments may be preferable for supporting the lower levels (Silvia 2012).

2.2.1.2. Challenging the taxonomy's directionality

Bloom's proposition that any idea must be *sufficiently understood* in order to be *applied productively* (Bloom and Krathwohl 1956) is plausible at least to the extent that deficits in low-level knowledge and comprehension are likely to propagate into the higher level processes, e.g., learners applying ideas incorrectly or not knowing how to apply them (or to analyse, evaluate, create, respectively). This seems plausible for all of the knowledge types considered in the revised taxonomy (factual, conceptual, procedural and metacognitive) and thus supports the suggestion that lectures should precede active learning, such as simulations (Silvia 2012).

This argument seems to contradict the popular stance that learning should start with hands-on exploration (Papert and Harel 1991a; Lipman 2003) and that active learning (e.g., using simulations) is useful in the service of - among other things - the lower-level learning objectives, i.e., comprehension (Moher 2006; Keller et al. 2007) and retention (Morgan 2003). Indeed, both positions have been supported in empirical studies with simulation-based teaching in formal education. This suggests that lower and higher level learning can mutually benefit each other (Figure 2.3).

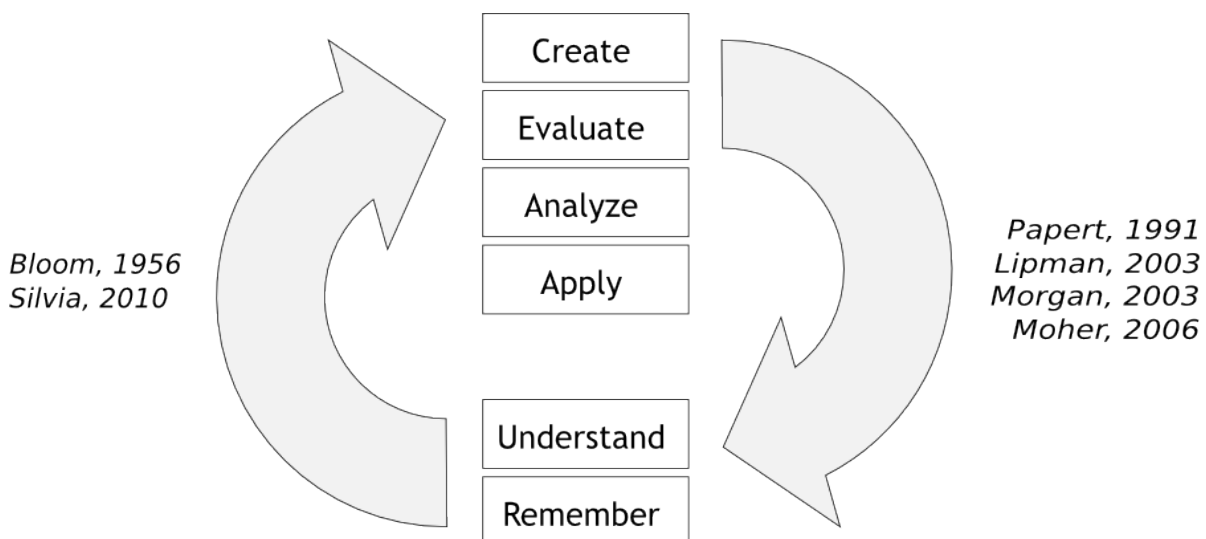


Figure 2.3.: Lower-level learning supports higher-level learning, and vice versa.

2.2.1.3. A circular interpretation

Considering the strong empirical evidence for active learning to support lower-level objectives (Morgan 2003; Moher 2006; Keller et al. 2007), must we dismiss Bloom's

opposing view, or can the two be somehow reconciled? The following paragraphs make an attempt to resolve the apparent conflict.

Arguably, Bloom's learning hierarchy applies to how a learner processes an individual *piece of knowledge* (a single fact, concept, etc.), whereas support for the opposite position comes from learners engaging with *networks of knowledge*, such as complex simulations. In the latter context it is plausible that having successfully applied one idea will help a learner grasp *related aspects* more easily. Therefore, perhaps the above diagram should be revised as shown in Figure 2.4 - with A and B representing either two related ideas (facts, processes, etc.) or two related aspects of the same idea:

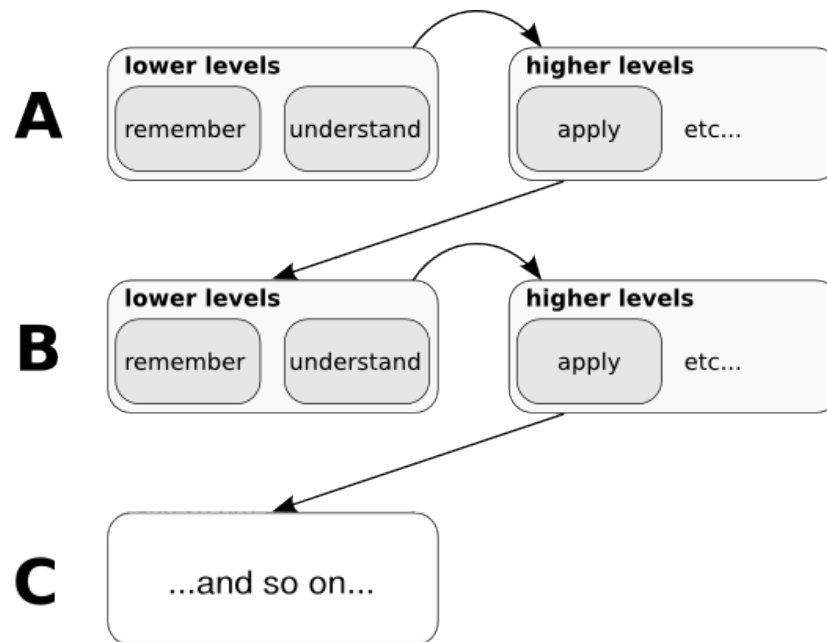


Figure 2.4.: An iterative variation of the above model, emphasising that applying one idea (A) helps learners understand a related idea (B).

This interpretation resolves the ostensible conflict and explains the success of active learning techniques for exploring complex knowledge networks, such as simulations (see Section 2.3.1.3). Furthermore, it supports a pedagogical approach of exploring knowledge domains in a cyclical fashion, by alternating between lower and higher level cognitive processes. This interpretation reconciles Bloom's key idea with other models of learning that stress the importance of iteration. This includes Kolb's experiential learning cycle (see Section 2.2.2.2 further below), which is a constructivist model and has explicit relevance to simulation-based learning research. This and other constructivist perspectives - including collaborative learning - are discussed below.

2.2.2. Constructivist theories

Constructivist theories have been instrumental to much significant research in active, collaborative and inquiry-based learning. In contrast to Bloom's taxonomy, which essentially derives from an analysis of practice in the formal education system, the following theories are mostly grounded in developmental psychology and cognitive science.

2.2.2.1. Overview

Based on the general hypothesis that knowledge and meaning are generated through experience with the world and other people, a number of ideas have been formulated that are fundamental to many contemporary approaches. These ideas include

- the idea that humans learn through experience and reflection on experience (Dewey 1933);
- the “zone of proximal development”, i.e., the difference between what learners can achieve in a collaborative situation versus on their own (Vygotsky 1978);
- “Internalisation”, i.e., social speech being transformed into inner speech (Vygotsky 1978);
- “Scaffolding” (Bruner 2009), i.e., contingent instruction, whereby a tutor gradually removes support in response to a learner becoming more autonomous. The premise is that responsibility for learning starts with the tutor and shifts gradually and at the right pace to the learner;
- Piaget's (1969) view that social interaction provides a catalyst for individual change.

Taking different angles on these basic ideas, a variety of approaches have focused specifically on *collaboration* as a way of learning. Three foundational theoretical positions were reviewed by Dillenbourg et al. (1996): socio-constructivist, socio-cultural and shared (or distributed) cognition. This list overlaps to some extent with the list that Stahl (2006) identified more recently, namely neo-Piagetian conflict theory, cultural-historical activity theory, social practice theory, Deweyan trans-actional inquiry and Bakhtinian dialogicality theory.

Although neither of these lists should be seen as definitive or exhaustive - even the labels may vary across discourses - together they provide a useful, rough overview of relevant movements in the context of CSCL. The list could be expanded with newer constructivist

concepts that emphasise peer dialogue, such as group cognition (Stahl 2006) and the social brain (Mercer 2013).

Constructivism is commonly contrasted with a prevalent stance in formal education that Jonassen (1991) calls “objectivism”. Similarly, constructivist teaching methods (e.g. discovery-based, problem-based, experiential or inquiry-based) are seen as contrary to “direct instruction” (Kirschner et al. 2006). The term *instructionism* was proposed by Papert to distinguish traditional teaching from his pedagogy of *constructionism*, which is based on constructivist principles (Papert and Harel 1991b).

2.2.2.2. Experiential learning: Kolb’s learning cycle

Kolb and Others (1984) conceptualises learning as a repeated cycle consisting of four parts: concrete experience, reflective observation, abstract conceptualisation and active experimentation (Figure 2.5).

The model has gained much popularity in many areas of instructional practice. A Google image search for “Kolb cycle” colourfully illustrates this through countless examples of the model being adapted and reinterpreted in different contexts. Its emphasis on active learning has made Kolb’s model one of the most frequently used models in the academic discourse on game-based and simulation-based learning, along with activity theory (Kuutti 1996) and constructivist theories (De Freitas and Oliver 2006; Lainema 2009).

Kolb’s model has a certain pragmatic vibe to it. More so than Bloom’s taxonomy, it proposes a specific rhythm in which to go about structuring learning-centred activities. Various flavours of the model can be found in guidelines and descriptions of professional practice today. Remarkably, this is not limited to purely *instructional* practice, such as professional training, but also includes design and engineering approaches that consider reflection and sense-making as instrumental parts of productive work.

It is worth noting that the cognitive process categories (i.e., plan, do, reflect, conceptualise) have been applied to individuals as well as to synchronised teams (Kayes et al. 2005). This emphasises the notion that groups can, to some extent, think and act collectively - quasi as a single cognitive agent engaged in a defined process of learning and productivity. Lainema (2009) identified experiential learning as one of the most influential approaches in the (somewhat industry-oriented) simulation and gaming literature.

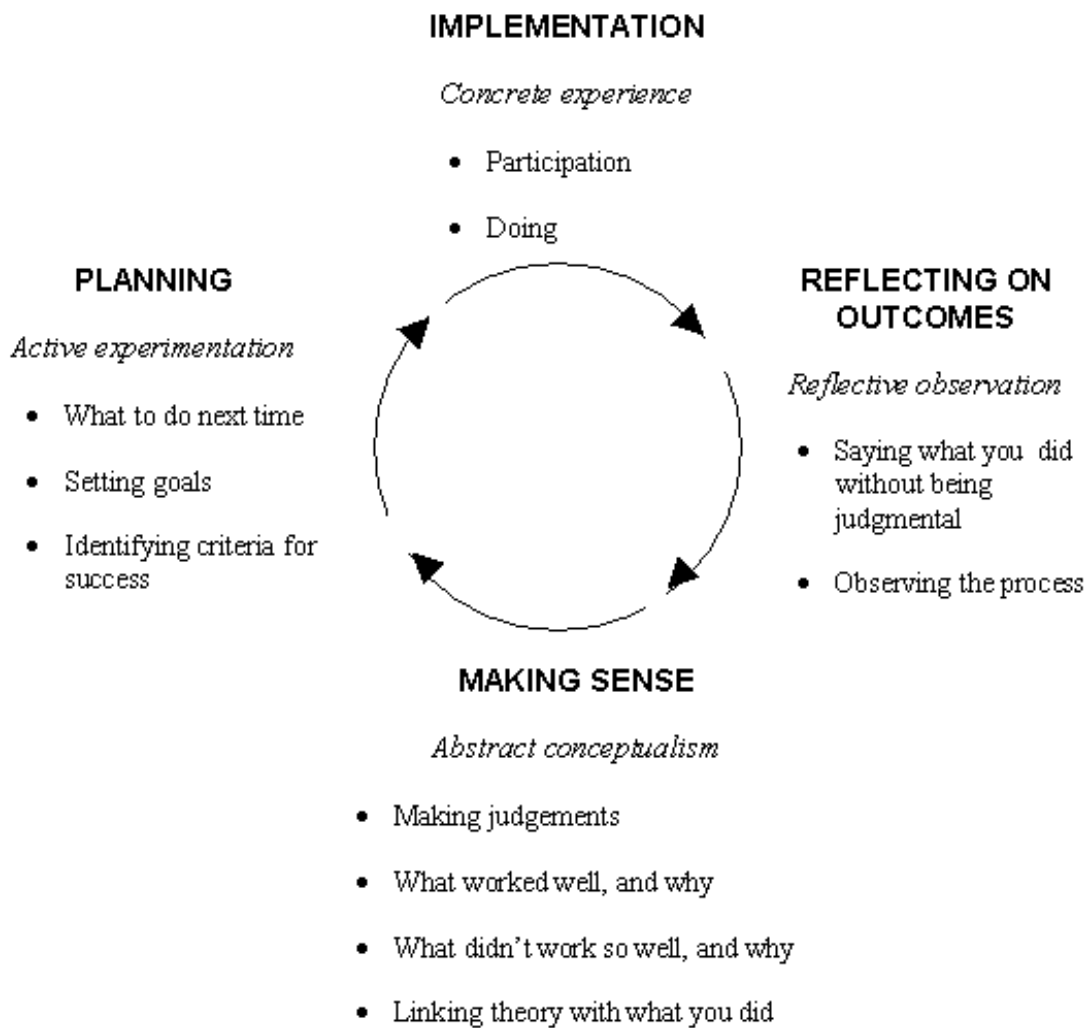


Figure 2.5.: Kolb learning cycle, extended by Cox et al. (2014)

Complementary to this top-down perspective on constructivist learning are other constructivist theories that focus on the complex, micro-level details of social interaction with peers and / or the mediating role of teachers and technology.

2.2.2.3. Learning through conflict and argumentation

Particularly relevant to CSCL are constructivists theories that emphasise the role of explicit verbal interaction between learners, and particularly the role of argumentation. Relevant psychological perspectives were reviewed by Nussbaum (2008), particularly sociocognitive conflict theory and cognitive elaboration theory. Nussbaum concludes that engaging in collaborative discourse and argumentation might have long-term effects in consolidating concrete learning gains. Jonassen and Kim (2010) agree with this view and further emphasise the general importance of argumentation as a problem-solving skill.

Much research has been concerned with assessing argumentation-based learning. Two broad categories are:

1. measuring the effects of argumentation on content retention and understanding, using pre and post tests;
2. assessing the quality of the constructed arguments themselves, e.g., how accurately students support their claims with evidence. A range of computer-supported methods have been proposed (Jonassen and Kim 2010). However, these methods require time-consuming post-hoc analysis of written arguments, making them impractical for daily teaching. There is a lack of formal methods for quickly assessing spoken, face-to-face argumentation on the fly, especially with regard to large groups.

2.2.2.4. Constructivism and teacher guidance

Although constructivism is a widely accepted stance among learning scholars, there are also critical voices. For instance, Kirschner et al. (2006) argue that discovery-based, problem-based, experiential, inquiry-based and constructivist teaching methods imply (quasi by definition) that students are given insufficient guidance, which necessarily results in inefficient learning. It is an ambitious claim that suffers - among other things - from an essential lack of clarity about what “guidance” entails. Short of a precise definition, the word appears in various meanings, including prior information and scaffolding during a task.

The apparent need for clarification in this matter is addressed in Dillenbourg and Jermann’s (2010) comprehensive account of a teacher’s guiding functions during classroom orchestration. The insinuation that active and collaborative learning are necessarily unguided, is strongly refuted by these authors. Nevertheless, the question whether teachers are generally able to deliver the appropriate kind and amount of guidance is still a legitimate concern, especially in the face of large group sizes, divided groups and real-world contingencies.

Another common criticism made against constructivist learning is that direct instruction is more time efficient. This argument, on the one hand, boils down to the uncontroversial observation that working out solutions tends to take more time than simply reading or being told the answers. On the other hand, proponents of direct instruction (including Kirschner et al., 2006) typically focus on lower-level learning, i.e., recall and comprehension, whilst disregarding higher-level learning gains as either irrelevant or impractical to measure.

2.3. Part two: Simulation-based learning

Simulations and role play have been used for teaching in many areas of education and training. Common reasons to use simulations instead of real settings involve cost, danger, inaccessibility and time (Rieber 1996).

The enormous variety of instructional goals, challenges, techniques, settings and learner populations is reflected in a diverse research landscape, with contributions from psychology, cognitive science, cultural studies, sociology, computing, economics, political science, education, and other disciplines. Several decades of research have produced knowledge that is potentially relevant to new designs, practice and research.

This section investigates pedagogical approaches based on collaborative simulations. Many of the cited studies follow a tradition of research and practice that is driven by concrete educational needs, rather than technical innovations or interfaces. Hence, the majority of best-practice knowledge in this area is predicated on old-school PC applications and non-digital (face-to-face) teaching of the kind that receives little attention in the contemporary HCI/CSCL literature.

However, the strong relevance of some of the work in this field to authentic professional practice is rarely found in HCI/CSCL studies. Therefore, one can find much potential for new insights by drawing new connections between these fields. This section surveys a number of common assumptions and approaches in simulation-based learning. A selection of these constitute the foundation for a newly devised pedagogical approach that is grounded in theory and applicable across topics and learner groups.

2.3.1. Common assumptions

There exists a wide range of assumptions regarding effective learning and teaching via role-play and computer simulations. Some of the most relevant assumptions are summarised in the table below (Table 2.1).

| Assumption | Type of support | Notable sources |
|--|-----------------|---|
| Simulations can engage a whole classroom-size group and encourage productive debate. | empirical | Wilensky & Stroup 1999; Colella 2000; Moher 2006; Keller et al., 2006 |
| Interactive simulations allow students to apply learning in practice. This shifts the emphasis onto higher-level learning, including learning to care. | empirical | Silvia, 2012 |
| Expert guidance (e.g. debriefing) is critical for students to draw accurate conclusions from a simulation. | theoretical | De Freitas & Oliver 2006; Fanning & Gaba 2007; Turkle 2003 |
| When guided by an expert, simulations can improve conceptual understanding. | empirical | Moher, 2005; Keller et al., 2006 |
| Simulations allow students to apply problem solving skills. | empirical | Shin et al., 2003 |
| Simulations allow students to develop complex problem solving skills. | theoretical | Gee, 2003 |
| Discussions based on simulations can foster skills development, such as argumentation, scientific thinking and empathy. | theoretical | Moher 2006; Wilensky 2006; Morgan, 2003 |
| Simulations help students develop empathy with a topic (e.g., a complex system) through first-person involvement | theoretical | Gee, 2006 |
| Building computer models may benefit learners' mental models of simulated phenomena. | theoretical | Jonassen & Strobel, 2006 |
| Through active learning, simulations can help students retain a higher percentage of what they learn. | empirical | Morgan, 2003 |
| Simulations can inspire and motivate students to take action in the real world. | theoretical | Morgan, 2003; Silvia, 2012 |
| Simulations are often considered too time-consuming to prepare and conduct. | survey | Faria and Wellington, 2004 |
| Actively modifying and creating simulations may be superior to readymade use in supporting deep understanding and literacy | theoretical | Turkle, 2003; Salen, 2007; Jonassen & Strobel, 2006 |
| Cognitive growth is "a dance between diving-in and stepping-out" (Ackermann, 1996) of a simulation. | theoretical | Resnick & Wilensky, 1997; Fanning & Gaba 2007 |

Table 2.1.: Summary of key assumptions regarding learning with simulations.

2.3.1.1. Where these assumptions come from

Many of these assumptions derive from areas of higher education and professional training where simulations have long traditions in practice and research (Fanning and Gaba 2007). Social aspects of learning, such as peer discussion, appear to have always been at the heart of these traditions (Morgan 2003). For instance, the Beer Game was developed in the 1960s at MIT as a supply chain simulation to help managers understand system dynamics (explained and revised by Kaminsky and Simchi-Levi 1998). Its effective use of small-group role-play and discussion has inspired voluminous research on simulation-based pedagogies (Meadows 1999), including participatory simulations (Wilensky and Stroup 1999), and remarkably it still ranks among the most widely adopted simulations in business education (Faria et al. 2009). Business education is looking back on half a century of experience with simulations (Faria et al. 2009) and over 30% of business professors were reported to be active simulation users (Faria and Wellington 2004).

By contrast, secondary education has a much shorter history of experience with discussion-oriented role-play simulations. Computer-supported pedagogies for schools date back as early as the 1980s, particularly the idea of *microworlds* (DiSessa 1977) and Seymour Papert's *Logo* programming language (Papert 1980), based on Papert's pedagogical idea of *constructionism* (Papert and Harel 1991a) which relies on constructivist theory. However, these approaches focused on a single child with a computer, rather than collaborative learning and affordances of the classroom.

It was not until the 1990s that Papert's ideas were extended to multi-learner interactions at classroom-scale. Notable developments include Resnick's (1996) proposal of *distributed constructionism* and StarLogo (Resnick 1997) and the first participatory simulations (Wilensky and Stroup 1999; Colella 2000) that were designed specifically for classrooms.

Although these whole-class activities were based on solid theories and unique learning benefits were indicated in small-scale evaluations, there has never been a large community of teachers employing these approaches on a regular basis. The same is true for the more recent approach of Embedded Phenomena (Moher 2006), which in some ways can be seen as a descendant of participatory simulations. Large-scale studies along the lines of Faria and Wellington (2004) to assess the adoption of simulations in everyday teaching practice have not been found for classroom contexts.

Finally, besides formal education, simulations and games have been studied with regard to learning in a wide range of fields, including computing, sociology, cultural studies and many others. For instance, James Paul Gee's theoretical works on video games

and simulations as high-level learning environments have been widely cited in discourses about developing literacies. Despite being mostly focused on individual learners and grounded in small-scale observations, his arguments about meaning making in complex problem spaces (e.g. Gee 2003) are too widely established to be ignored. Other authors (e.g. Turkle 2003) have also contributed relevant insights, based on cases from consumer video games.

Several academic venues are dedicated to formal and informal learning with simulations, video games and Serious Games, such as the European Conference on Games-Based Learning and the journal of Simulation & Gaming. The above table represents only a relevant fraction of fairly well-supported claims in these areas.

2.3.1.2. Assumptions about benefits

The following subsections elaborate the above assumptions in more detail, starting with the general benefits of simulations for learning and continuing with assumptions about how to use simulations effectively in teaching practice.

a) Active learning with peers

The notion that discussion with peers can benefit learning is commonly accepted among constructivist scholars and is particularly held high in the context of simulations. For example:

In arguing for the effectiveness of role-playing activities in International Relations courses, (Morgan 2003) cites a study indicating that “active leaning” (notably group discussion and teaching others) leads to better retention rates than “passive” modes of delivery (Fig. 2.6).

Stahl’s (2006) theory of group cognition is grounded in empirical studies focusing on peer discussion. Some of these studies used simulations as the object of debate.

b) Developing argumentation skills

Several simulation-based pedagogies have aimed to help students develop scientific thinking and argumentation, by emphasising shared hypothesising and discussion. These include participatory simulations (Wilensky and Stroup 1999; Colella 2000), Embedded

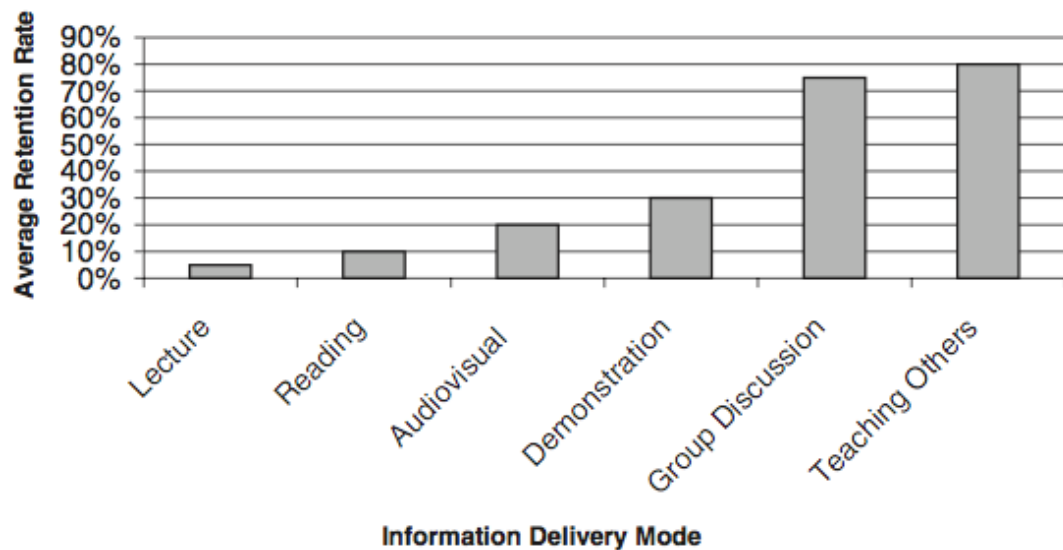


Figure 2.6.: Active learning is considered superior in helping students retain their knowledge over time (from Morgan, 2003).

Phenomena (Moher 2006), scientific discovery learning more generally (De Jong and Van Joolingen 1998) as well as approaches that allow students to create their own simulations (Resnick 1996; Wilensky and Reisman 2006).

A common argument among these authors is that engaging and supporting students in authentic scientific discourse is likely to develop their scientific thinking and argumentation skills in the long term.

The argument is well supported by various learning theories, although empirical evidence has been lacking, according to a recent literature review that also considered various studies on participatory simulations (Hilton and Honey 2011, p.36). To explain this lack of evidence, a number of reasons are conceivable: Firstly, scientific argumentation comprises a variety of high-level cognitive skills that are difficult to quantify and measure in an individual or a group. Secondly, such skills are known to require a long time period to develop. Finally, exposing a large sample of groups to simulation-based learning over such long periods is a challenging undertaking, due to a multitude of obstacles (including those outlined in Section 2.4.3).

Nevertheless, the hypothesis that argumentation skills are likely to increase through practice is reasonable and has not been disproved. Moreover, the review endorsed empirical evidence suggesting that simulations can improve conceptual understanding by engaging students in productive discourse (Keller et al. 2007) and scaffolding students' construction of scientific arguments (Sandoval and Reiser 2004).

c) Developing problem-solving skills

Constructivist scholars have argued that interactive simulations provide opportunities for applying problem solving skills, including well-structured and ill-structured problems (Shin et al. 2003). Gee (2008) emphasises the importance of situated understandings that simulations can provide, in addition to what he calls general or verbal understandings:

“A general or verbal understanding implies an ability to explicate one’s understanding in terms of other words or general principles, but not necessarily an ability to apply this knowledge to actual situations. Thus, while verbal or general understandings may facilitate passing certain sorts of information-focused tests, they do not necessarily facilitate actual problem solving.” (Gee 2008)

d) Developing empathy

Empathy is frequently discussed as a learning goal in the context of simulations and role-play (De Freitas and Oliver 2006; Fanning and Gaba 2007; Flanagan 2009; Krain and Lantis 2006; Whitton 2009). Gee (2007) stresses that players of a simulation can develop empathy with a complex system by picturing themselves inside it.

With debate-oriented role-play activities it is common that participants are divided into groups that represent populations with different interests. For instance, in the context of teaching international relations, Morgan (2003) cites a simulation that replicates a political crisis between Athens and Melos around 415-421 BC. Participants initially take one side or the other, to the effect that they will be primed with different assumptions - in this case - realist versus idealist perspectives on common issues of international relations. The author argues that such simulations - by revealing contending perceptions of reality and supporting mutual empathy - can highlight (a) why conflict is so prevalent and (b) how difficult global cooperation can be to achieve.

e) Supporting high-level and low-level learning

While much of the literature focuses on the unique potential of simulations to support high-level learning, such as applying concepts in realistic situations (Gee 2003; Silvia 2012), some empirical studies have also measured increases in students’ low-level comprehension and retention of conceptual and procedural knowledge (e.g. Keller et al. 2007; Moher et al. 2005).

2.3.1.3. Assumptions about effective use

It is clear that simulations are just tools. They need to be used wisely in order to support the desired goals and to avoid adverse effects (such as confusion, false conclusions, etc.). Below I elaborate on a number of caveats that have been pointed out by various authors.

a) Glass box better than black box?

Jonassen and Strobel (2006) distinguish two types of simulation-based tools for learning: One type is “*black box systems*” that allow users to infer parts of a predefined model by observing how the outputs change for different inputs, while the underlying transfer functions (or rules) are hidden from the user. The other type, referred to as “*glass box systems*”, allow learners themselves to look inside the inner workings of the model (Figure 2.7). The latter category includes systems that allow users to construct models from scratch - or at least make changes to existing models at a deep and meaningful level.



Figure 2.7.: The rules are known in a glass box system (left), whereas in a black box system they are not (right).

Jonassen and Strobel (2006) mention a lack of research investigating the cognitive development-related differences between building models versus using existing models. However, it seems widely understood that a poorly reflected, unsupervised use of black-box simulations can be harmful.

For instance, Turkle (2003) describes how mastering the game of SimCity led a teenager to the false conclusion that raising taxes leads to riots in a city. Turkle’s explanation was that the SimCity game, being a black-box simulation, failed to encourage users to think deeply and critically about the underlying rules of the simulation. Consequently, Turkle argues for the importance of people becoming literate in simulations as a medium - ideally including learning how to program simulations oneself - and developing a “critical stance”. The responsibilities of students and educators are discussed in the context of university education.

While Turkle's argument culminates in endorsing digital literacy, i.e., a generally more empowered culture of dealing with computing technology (including black box systems), Jonassen and Strobel's argument for using glass-box modelling tools focuses on learners' development of mental models of concrete domains. Both arguments outline desirable goals for simulation-based teaching. Moreover, both arguments provide some theoretical support for the benefits of simulations that are glass boxes, rather than black boxes.

Nevertheless, the black box paradigm is prevalent in many classroom-oriented approaches that emphasise discovery and hypothesising. Often, the models that students are given to explore are designed to be *opaque*, i.e., the rules are deliberately hidden, in order to allow for hypothesising. This applies, for instance, to the majority of "scientific discovery learning", whereby "*the main task of the learner [is] to infer, through experimentation, characteristics of the model underlying the simulation*" (De Jong and Van Joolingen 1998). Part of this genre are many microworlds, participatory simulations and embedded phenomena. Notable classroom studies in these areas have emphasised the teacher's (or instructional designer's) responsibility to scaffold students' reasoning during or after the activity (De Jong and Van Joolingen 1998; Colella 2000; Moher 2006).

b) Stepping out and debriefing

In a participatory simulation context, Colella (2000) emphasised that "considerable learning occurred as students were able to step back from their immediate experience and analyze the situation". The author refers to Ackermann's (1996) notion of "diving-in" and "stepping-out", i.e., students moving back and forth between full immersion in a problem and thinking about a problem. The pedagogical value of this "back and forth" is argued convincingly, although little detail was provided regarding how the teacher and technology were involved in orchestrating this fluidly.

The common practice of "debriefing" after a simulation can be seen as a particular method of stepping out, whereby participants are guided by an instructor. Debriefing can take various forms, collective or one-on-one, and involve different media, e.g., lecture, feedback, discussion or video playback (Savoldelli et al. 2006).

The importance of debriefing has been widely emphasised in research on professional training, particularly with regard to dangerous fields such as medicine, military, emergency, etc. In the medical context Fanning and Gaba (2007) critically review established approaches to simulation debriefing and discuss typical goals, methods and styles. Typical goals include helping participants process the experience cognitively and emotionally and

make accurate generalisations when comparing to real-life situations. Since those goals and some of the methods discussed apply universally to simulation-based teaching, they were taken into account in the design and evaluation of simulations in this research.

Debriefing is also mentioned as one of three important functions of lecturing in Dillenbourg & Jermann's (2010) orchestration paper. The other two functions concern giving students the right vocabulary and metacognitive context to engage with a topic effectively.

c) Supporting higher-level learning

The potential to engage learners and support conceptual understanding can make simulations appealing as introductory activities to a topic. However, in some cases constructivist activities (such as simulations) may be more effective when students already have some basic understanding of the topic to begin with (Kirschner et al. 2006; Silvia 2012).

Silvia (2012) argues that traditional pedagogical techniques - such as reading assignments, lectures, tests and papers - are well suited for establishing the basics, i.e., helping students acquire theoretical concepts. Developing students' ability to *apply* concepts to realistic situations - i.e., higher level learning - is considered to be the point where simulations can really add value.

Silvia's argument is based on Bloom's (1956) original version of the taxonomy that stresses the hierarchical nature of the six levels: i.e., in order to "apply", one must first "understand", and so on. A more recent, revised version of the taxonomy is more relaxed in these terms (Krathwohl 2002), suggesting that the categories overlap considerably and teachers have some freedom to use them differently. Indeed, as Wegerif (2006) notes, some authors, such as Lipman (2003) propose that one should teach in exactly the opposite direction, starting with the most high-level kind of dialog, in the belief that all the dependencies are more or less automatically sorted out in the process.

Silvia's argument is nevertheless compelling. It raises the question of how simulation and traditional teaching can complement each other.

2.3.2. Relevant approaches

Simulation-based pedagogies differ in many respects. Two salient criteria in this context are:

1. whether participants *observe* the situation from an objective point of view or engage in *role-play*, i.e., playing a role as an agent within the simulation;
2. to what extent users can see and change the rules that underly the model. This extent may vary from tweaking a limit set of exposed input variables to essentially changing the model's source code.

Figure 2.8 represents these two distinctions as a 2x2 table with some examples. All the examples are of computing-supported approaches, although in principle the categories would also apply to non-digital simulations, such as board games, card games or paper-based role-play.

| | <i>role-play</i> | <i>observation</i> |
|------------------|--|-----------------------------------|
| <i>glass box</i> | game design / modding virtual worlds | modelling |
| <i>black box</i> | serious games participatory simulations | microworlds embedded phenomena |

Figure 2.8.: Some simulation-based pedagogies, roughly categorised.

This categorisation illustrates the breadth of common approaches from a perspective that is relevant to this dissertation. It is neither exhaustive nor definitive, i.e., the categories represent *typical characteristics* of each approach, rather than necessary or sufficient criteria. For instance, neither participatory simulations nor embedded phenomena need necessarily be black box systems - although so far the majority of published applications have been, as they did not allow students to see or change the underlying rules.¹ Similarly, several applications that go under the umbrella term “microworld” (or “serious game”, respectively) could be categorised differently and practice is full of grey areas. The purpose of the above diagram is not to provide a full taxonomy of concrete approaches

¹Some attempts were made in this dissertation to add those features to a participatory simulation, aiming to devise a more glass-box like approach, see Chapter 6.

but rather to sensitise the readers to those four quadrants, each of which represents a different focus in instructional design. Discussing the rationales behind these different foci - and their pedagogical ramifications - is the goal of the following paragraphs.

Surprisingly or not, the categorisation revealed that for almost each role-play-based approach (left column in Figure 2.8) there exists a somewhat comparable observation-based approach (right column). Below I will elaborate on each approach. Starting from the top left and proceeding in rows downwards through the figure, I will highlight some significant relationships between related approaches. Special attention will be given to those apparent horizontal pairs, as they help illustrate the pedagogical implications of role-play and observation-based engagement with simulations. One of these pairs - namely participatory simulations and embedded phenomena - will be elaborated in particular detail, since those two approaches focus on large groups of participants.

2.3.2.1. Glass box approaches

The following approaches emphasise learning by creating or modifying simulations.

a) Game design and game modification

In entry-level computing education, students are often taught how to build games and simulations, on the premise that many novice students find this engaging, while it also introduces them to fundamental concepts and skills, such as object-oriented development. Based on this premise, some school and university teachers have employed beginner-friendly programming tools, such as GameMaker (Overmars 2004), AgentSheets (Repenning and Ioannidou 2008), Alice, Greenfoot and Scratch (Utting et al. 2010).

Scaffolding learners' transition from passive consumers to active producers is a common theme in those pedagogies. As an intermediate step midway through this journey, several pedagogies have encouraged learners to make informed modifications to existing simulations and games. Most programming tools enable this as a standard feature implicit in the ability to open and save documents. Moreover, some applications and pedagogies are explicitly geared towards modifying, sharing and remixing games and simulations that others have created. Examples include the Scratch online community (Monroy-Hernández and Resnick 2008) and Gamestar Mechanic, an online game designed to teach players the fundamentals of game design through "fixing" examples of mechanically flawed games (Salen 2007). These approaches emphasise higher learning goals beyond the immediate

technical skills, such as students' development of reflective design practice, digital literacy and meaningful participation in knowledge communities (Salen 2007; Kafai et al. 2009).

Some of those community and literacy-oriented approaches have looked at learning through modifying off-the-shelf computer games. Relevant concepts are:

- *modding*, i.e., creatively hacking a game's content (e.g., images, sounds, narrative) to create new meanings (Steinkuehler and Johnson 2009);
- *soft modding*, i.e., informally defining new challenges for other players, without actual hacking (Hayes and Gee 2010); and
- *subversive play*, i.e., playing in deliberate contrast to the values promoted by the game designers (Flanagan 2009);

Evidence for literacy learning through commercial games has mostly derived from settings that are less constrained than classrooms. For instance, Squire and Barab (2004) has observed young people in after-school clubs playing Civilization III for 18-20 hours. Hayes and Gee (2010) examined online forum discussions and stories from spare-time players. By contrast, within the time, curriculum and assessment constraints in formal education, targeting these learning goals at scale would require more focused solutions (Egenfeldt-Nielsen 2006; Rice 2009).

Moreover, the challenge to find efficient ways of realising game making, game modding and programming activities in a collaborative, large-group context with adequate teacher support remains an unsolved problem for research and instructional design.

b) Modelling

A similar range of approaches can be summarised under the umbrella term of "modelling". There is much overlap between game design and modelling with regard to primary learning goals and tools for authoring simulations (e.g., Scratch, StarLogo TNG). A key difference can be seen in how the modelling approach tends to focus on objective descriptions of phenomena, whereas game design also emphasises subjective player experience.

Jonassen and Strobel (2006) provide a review of fundamental theories and a classification of relevant tools that illustrate the breadth of pedagogical perspectives. A relevant distinction is made between two categories of tools: "black box systems" and "glass box systems" (see Section 2.3.1.3) The authors explicitly situate most microworlds and simulations in the "black box" category, as those systems allow learners merely to see and

manipulate a superficial subset of variables. By contrast, the “model building” family of approaches discussed here deals with “glass box systems”, i.e., desktop software that allows learners to model (thus articulate) how they think about a certain domain. The authors apply this idea of mental models to individual cognition as well as collaborative, group mental models. Four types of knowledge are proposed to support the idea of collaborative mental models:

- activity-based knowledge
- conversational/discourse knowledge
- social/relational knowledge
- artifactual knowledge , i.e., “cognitive residue evidenced in the artifacts that learners produce”.

Jonassen & Strobel’s (2006) account explicitly includes simulation-authoring tools (e.g., Stella, PowerSim, VenSim, and Model-It) as well as expert systems and semantic modelling tools (e.g., concept maps). Their focus on mental models and argumentation (Jonassen and Kim 2010) is compatible with studies in science education that encouraged learners to build and discuss models of natural phenomena. Examples include an increasing body of work using the *Logo* family of languages, e.g., using NetLogo to help students articulate and critique theories of predator-prey relationships (Wilensky and Reisman 2006).

c) Virtual Worlds

Virtual worlds, such as Second Life, are multi-user, sandbox environments that enable various forms of real-time collaboration and model building. They can be described as glass box systems to the extent that users can build “anything” within the liberal confines of the system, using accessible graphical and scripting interfaces. For instance, the S4SL middleware allows Scratch programming within Second Life.

The ability to integrate powerful programming code into a virtual world makes this type of software potentially interesting for designing collaborative problem-solving activities based on avatars in 3D space, e.g. for team building (Inman et al. 2010). However, the use of proprietary virtual worlds is prone to a range of technical problems that jeopardise a consistently smooth user experience. These include unpredictable restrictions on the

ability to create and manipulate in-world objects, server errors, delays and automatic software updates (Ellis et al. 2008).

Orchestrating classroom-based use of virtual worlds in classrooms is difficult due to several issues, including:

- the usual limitations of desktop-based (single user) interfaces for large-group collaboration;
- complex user interfaces likely requiring frequent teacher assistance.
- the effort it takes to learn and perform basic operations, such as walking around, camera control, interacting with objects and avatars, etc., - taking time and cognitive resources off the actual learning;
- the potential for learners to get distracted and fool around with the avatars;

2.3.2.2. Black box approaches

Another set of approaches is centred around models that learners can only partially infer and manipulate. This typically includes serious games, microworlds, participatory simulations and a range of science-inquiry approaches, such as embedded phenomena.

a) Serious games

Under the label of “serious games” an emergent thread of research has been concerned with using, designing and using computer games specifically with instructional intentions in mind. Fundamental to this approach is the idea to leverage experience from computer games to design engaging, constructivist learning environments.

Unlike other simulation-based approaches, serious games tend to emphasise various mechanisms of in-game scaffolding, such as incremental challenge (Kelly et al. 2007) and embedded assessment (Ritterfeld et al. 2010) to support the player on a journey through a series of learning objectives.

Figure 2.8 situates this approach in the “role play” category, based on the premise that players see themselves as being ‘inside’ the complex system (Gee 2007). This cognitive interpretation of a “first person perspective” is not affected by the fact that graphical representations may be first person or third person.

Coming from a video games-inspired tradition, many serious games have focused on single players progressing at their own, individual pace. While this feature can benefit individual learning, integrating such single-player experiences into shared classroom contexts is problematic (Egenfeldt-Nielsen 2004). In orchestration terms, the problems arising concern limitations of the teacher's awareness, linearity and cross-plane integration (Dillenbourg and Jermann 2010).

b) Microworlds

Microworlds are simulated environments for individuals (children) to learn about abstract ideas by manipulating concrete, screen-based representations and exploring the effects of those manipulations (e.g. DiSessa 1977; Papert 1980; Hoyles et al. 2002). The key pedagogical idea is based on the Piagetian view that facility with concrete representations develops at an earlier stage than facility with abstract representation (Roschelle and Pea 2002).

The design principle of making things easy for learners to grasp is also highlighted by Rieber (1996) in distinguishing microworlds from simulations, based on two criteria:

1. *"First, a microworld presents the learner with the "simplest case" of the domain, even though the learner would usually be given the means to reshape the microworld to explore increasingly more sophisticated and complex ideas."*
2. *"Second, a microworld must match the learner's cognitive and affective state. Learners immediately know what to do with a microworld - little or no training is necessary to begin using it. [...] In contrast, a simulation is determined by the content or domain it seeks to model [...] For example, most flight simulators would not be considered microworlds for most people because they would be quickly overwhelmed with the environment."*

While microworlds transform abstract ideas into concrete sign systems (on the screen), participatory simulations take this principle one step further, resulting in embodied, physical, spatial explorations that precede concrete sign systems (Roschelle and Pea 2002).

c) Participatory Simulations

The participatory simulations approach emphasises that students become *first-person* participants in the simulation via role-play (e.g. Wilensky and Stroup 1999; Colella 2000). In contrast to most microworlds, this approach aims to liberate participants from the limitations of personal computer screens:

“Participatory simulations build on the characteristics of microworlds, in which models can be executed, and augment them with the affordances of real-world experience.” (Colella 2000)

The notion of “players” is characteristic for participatory simulations, in the sense that participants play an active part (not just observing), much like players in a computer game - except that the emphasis is on face-to-face communication, rather than limiting interaction to the screen. For this purpose, students are equipped with computing devices, such as mobile handhelds (e.g. Wilensky and Stroup 1999; Klopfer et al. 2002 2005) or wearable devices (Colella 2000) through which each participant controls a small part of the simulation.

For example, a classic study allowed students to explore how virus diseases spread in a crowd of people (Colella 2000). Students moved around in the classroom, wearing tiny networked computers (‘Thinking Tags’) that tracked their encounters with other students in the room.

“Participants become players in a computationally mediated system comprised of people and their [...] personal computers.” (Colella 2000)

In addition to personal devices, some participatory simulations featured additional public displays for the emergent results to be seen by all participants and discussed with the teacher (e.g. Wilensky and Stroup 2000).

Moreover, some participatory simulations have involved elements of *augmented reality* to make the experience more visceral and immersive for learners. For instance, the “Environmental Detectives” simulation (Klopfer and Squire 2008) mapped a simulation of an environmental disaster onto a real, physical watershed with streams, trees, and other natural elements. Using mobile handheld devices as “windows into the simulation”, students had to collect and combine data from both the real and the virtual in order to find the cause of the problem. Another outdoors example is “Savannah”, where groups of students had to survive as a pack of lions in a virtual savannah (in reality an open field of grass), again using mobiles (Benford et al. 2004).

d) Embedded Phenomena

Tom Moher's approach of Embedded Phenomena (Moher 2006) was already mentioned, which generally involves classroom-size, augmented reality simulations of natural phenomena that students can observe and analyse together.

The approach was inspired by participatory simulations and shares a number of typical characteristics, including:

- a focus on multi-user, face-to-face activity that involves the entire classroom, with students looking and moving in all directions;
- fixed, black-box models;
- learning goals that typically revolve around generating and testing hypotheses.

Embedded phenomena are distinguished from participatory simulations by a number of characteristic features, typically including the following:

1. Students take part as third-person observers who can analyse, but not influence the simulated phenomena.
2. All the computing devices and low-tech materials are usually *shared* between all participants, rather than each participant (or small group) being equipped with their own device.
3. The physical space of the classroom represents the physical space of the simulation, e.g., an earthquake area (Moher et al. 2005) or creatures crawling around inside the classroom walls among pipes and wiring (Moher et al. 2008). This direct mapping immerses students both physically and virtually. Arguably, this feature is necessary for the embedded phenomena approach, whereas participatory simulations may have it or not.

Participatory simulations and embedded phenomena are the two main approaches that truly capitalise on the affordances of a classroom as a large-group, co-located learning space with teacher support, in contrast to the majority of computer simulations that rely on single-user interaction with a screen.

2.3.2.3. Discussion of relevant approaches and trends

Differentiating between (a) role-play and observation and (b) black box and glass box approaches leads to a number of relevant questions and working hypotheses.

a) Does whole-classroom imply black box?

So far we have identified two simulation-based approaches that involve the whole classroom: (a) participatory simulations and (b) embedded phenomena. In both cases, the focus has been on discovery learning, which is a black box paradigm by design: The models that students are given to explore are deliberately *opaque*, i.e., the rules are hidden, to allow for hypothesising.

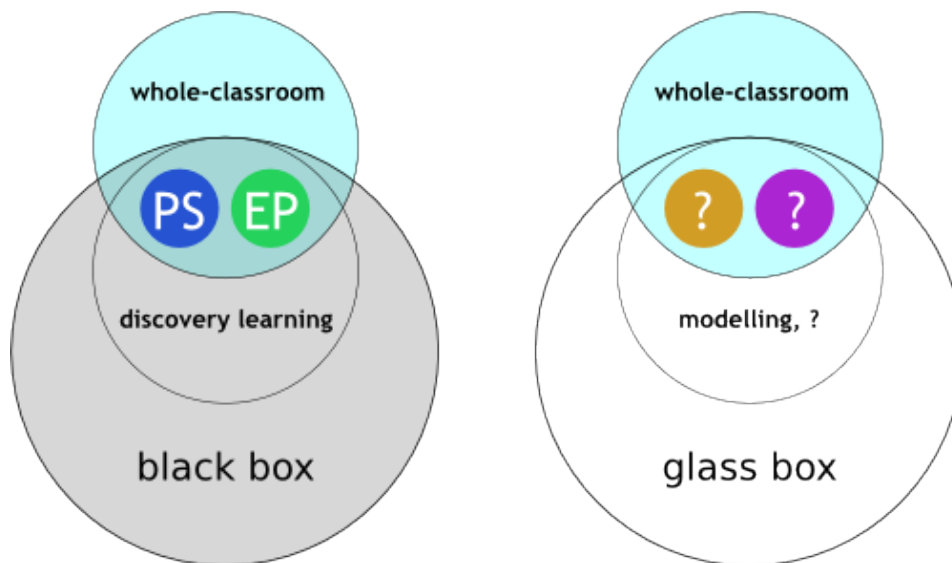


Figure 2.9.: In the whole-classroom context, participatory simulations (PS) and embedded phenomena (EP) belong to the black box family (left). It is unclear what a glass box equivalent might be (right).

By contrast, glass box approaches appear to be rooted in a tradition of personal computing, making them difficult to orchestrate well in classrooms. This raises an important question: Could the mechanisms that enable whole-class collaboration be leveraged to support other glass box approaches, such as modelling?

Surprisingly, in practice there has been little or no overlap between (a) modelling and (b) whole-classroom simulations. This is even more surprising considering that some researchers have pursued both directions. For instance, Resnick & Wilensky's (1998) StarLogo-based group activities notably informed work on participatory simulations (e.g.

Wilensky and Stroup 1999; Klopfer and Squire 2008) as well as PC-based modelling approaches using StarLogo-descendant tools, such as NetLogo (Wilensky and Reisman 2006) and StarLogo TNG (Klopfer et al. 2009).

In the absence of modelling software that is specifically designed for whole-classroom use, it is currently not obvious how these two approaches could be reconciled to combine their benefits.

b) What about participatory modelling?

Outside the education context, participatory design of (computer) models and simulations is a common approach. For instance, many projects to engage local stakeholders in face-to-face dialogue about environmental, social or political issues have combined modelling techniques with role-play elements and novice-friendly interfaces.

Different labels exist for various flavours of this approach, including “participatory modelling” (Andersson et al. 2008) and “companion modelling” (Barreteau 2003). Even the term “participatory simulation” has been proposed in this related context (Castella et al. 2005; Briot et al. 2007). However, there appears to be no explicit connection between this approach and any of the classroom-based studies considered here.

Typical studies in the “participatory modelling” category appear to focus on expert-guided, small groups - using PC screens (e.g. Briot et al. 2007) and/or low-tech materials like paper and board games (Castella et al. 2005) - rather than classroom scale. The challenge of taking modelling to the whole-classroom scale appears to be an unsolved problem.

c) Communities of practice

Several approaches have emphasised the idea of learning by participating in a community of practice (Lave and Wenger 1991). This concept has been applied to different types of communities at different scales. One end of the spectrum considers massive, remote communities of learners creating and exchanging simulations, such as the Scratch online community (Monroy-Hernández and Resnick 2008). On the other end of the spectrum, studies have focused on a classroom full of students learning together as a community, while exploring one and the same simulation face to face (Moher 2006). The latter approach has been linked to the related “knowledge community approach” in support of inquiry-oriented science curricula (Cober et al. 2012; Lui et al. 2014).

d) Attention to classroom orchestration

It appears that many of the relevant simulation-based learning studies have lacked attention to classroom orchestration issues - although arguably recent trends suggest a turn towards orchestration (Arnab et al. 2012). An exception seems to be Tom Moher's group whose designs have always taken real-life classroom constraints seriously from the beginning.

In the first paper introducing embedded phenomena, Moher et al. (2005) already hints at the difficulties some teachers might have in organising a large group in physical motion. This was well before the remarkable boom of "orchestration" as a popular term in the learning literature (Prieto 2012, p.52). Since the group has continued to produce new designs within the embedded phenomena approach (e.g. Novellis and Moher 2011; Gnoli et al. 2014) and inspired a wider range of research on classroom-scale simulations and visualisations as immersive learning spaces (e.g. Slotta 2010; Smørdal et al. 2012; Lui et al. 2014).

2.3.3. First outcome: A lean pedagogy for applying theory in practice

To summarise this section, I conclude that simulations offer a great variety of benefits for learning. Particularly, they allow students to apply theory in practice through discussion, collaboration and problem solving. However, certain factors should be in place in order to realise those benefits and avoid adverse effects, such as false conclusions. These important factors include guidance by a human expert, debriefing, an adequate rhythm of diving-in and stepping-out and the ability for students to inspect the rules of the simulation and perhaps even change those rules.

The pedagogical approach adopted in this thesis aims to account for these factors by carefully integrating lectures, simulation and discussion. Many of the common assumptions and practices discussed above are embodied in this approach, as illustrated schematically in Fig. 2.10 and elaborated below.

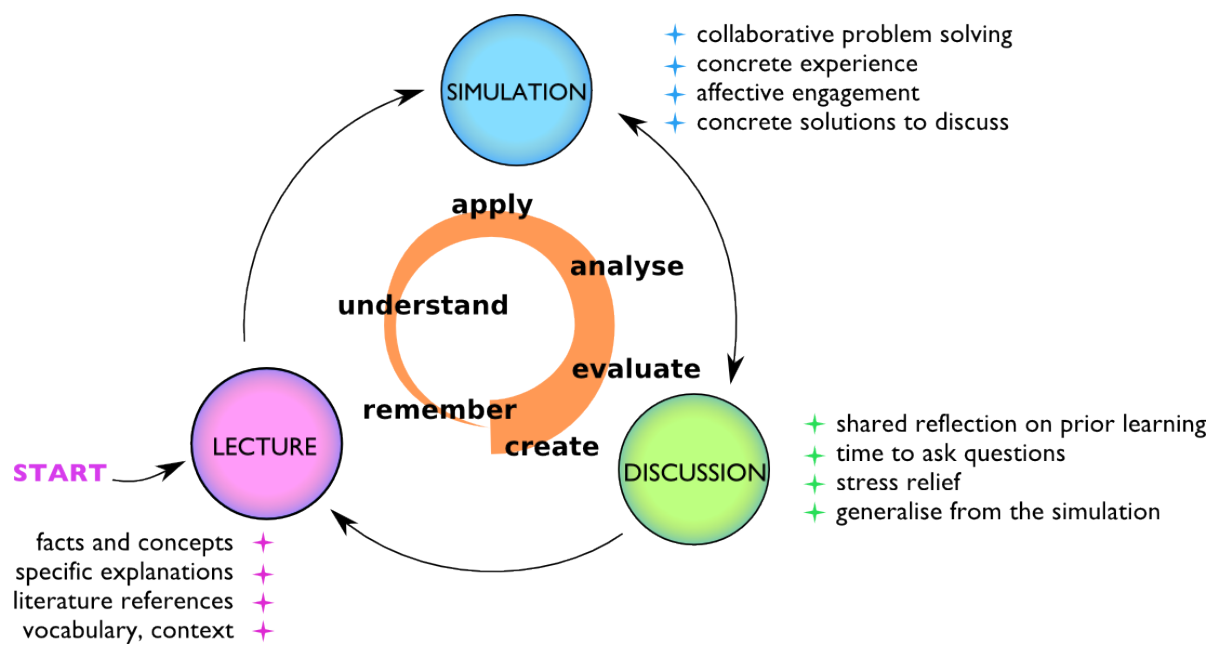


Figure 2.10.: Repeated cycles of lecture, simulation and discussion are proposed to support a continuum of low-level and high-level learning objectives.

The figure highlights a number of core principles:

1. The three teaching methods (lecture, simulation and discussion) are associated by proximity with the six categories from the revised version of Bloom's taxonomy (the spiral in the middle) and some examples of concrete learning objectives specific to the approach (bullet points along the outside).
2. There are no strict borders between the categories but the proximities indicate closer association. Whilst lectures and reading assignments are deemed efficient for helping students remember and understand facts and concepts (i.e., lower-level learning) simulation-based role-play and discussion are proposed as a complementary strategy.
3. The arrows indicate that it is recommended to start with lectures, then proceed to simulation, followed by discussion. As needed, discussion may be followed by further lectures or further rounds of simulation, as indicated by the bidirectional arrow on the right.
4. Going through the whole cycle is assumed to address the entire panorama from lower to higher level learning objectives, roughly in the sequence originally suggested by Bloom and Krathwohl (1956) and supported by (Silvia 2012).
5. In analogy to Kolb's learning cycle (1984), repeated iterations of the cycle are suggested to integrate the benefits of abstract conceptualisation and concrete experience.

The primary learning objectives are for all students to apply and analyse their lecture-based understandings in shared problem solving, supported by distributed computer simulations. Moreover, by evaluating their solutions with the teacher - which may lead to creative modifications to the simulation - students reach the high end of Bloom's continuum.

This description of a pedagogical approach constitutes the first outcome of the literature review. It aims to provide a starting point for instructional design, particularly how to integrate an existing simulation within a lesson, by outlining a set of instructional goals. However, at this stage the description contains no helpful information on how to design towards these goals. Particularly, it says nothing about the role of technology, the relationship between the teacher and the technology, which types of technology and interaction techniques to use, etc. These questions are further investigated in the following sections.

2.4. Part three: Computer-Supported Collaborative Learning

This section reviews relevant literature in the field of Computer-Supported Collaborative Learning (CSCL), a branch of the learning sciences concerned with studying how people can learn together with the help of computers (Stahl et al. 2006). Since the early 1990s this field has accumulated a voluminous body of conceptual, empirical and design knowledge that is relevant to classroom teaching and has the potential to inform new designs.

This part of the review focuses on the following questions:

1. Which of the learning objectives identified above are realistically feasible within the constraints of a classroom setting?
2. What are the realistic capacities of a single teacher to in managing a technology-enhanced, whole-class activity?
3. How can technology and a human teacher complement each other most efficiently in this context?

After a general overview, a number of particularly relevant approaches are discussed in detail. The focus lies on the recently developed concept of *classroom orchestration* and how insights from small-group interaction can be practically extended to large groups. The conclusions drawn in this section support the suggestion that design should begin by considering *the whole classroom as the user* (Dillenbourg et al. 2011), rather than attempting to scale up small-group tasks.

2.4.1. Trends in the field

In retracing the history of the field, Stahl et al. (2006) identify a number of significant turns describing high-level shift of focus in CSCL:

1. from artificial intelligence to collaboration support
2. from individuals to interacting groups
3. from mental representations to interactional meaning making
4. from quantitative comparisons to micro case studies

Points 2 and 3 are highly relevant and discussed below. Point 4 is discussed in Chapter 3.

2.4.1.1. From individual cognition to interacting groups... and orchestration

Arguably, the field's increasing attention towards whole-class scenarios has been motivated by two main factors:

On the one hand, there has been notable support from underlying disciplines, such as cognitive science and social psychology, advancing more socially-oriented theories, holding that intelligence does not reside within the individual alone. For instance, Dillenbourg et al. (1996) emphasised the influence of situated cognition (Suchman 1987; Lave 1988) and distributed cognition (Hutchins 1991). These concepts have also remarkably shaped other areas that rely on cognitive science and psychology, such as recent perspectives in Human-Computer Interaction (Rogers 2012).

On the other hand, an increasing number of CSCL researchers have focused on practical solutions to real-life learning requirements in classrooms (Dillenbourg et al. 2012).

The recent turn to orchestration can be seen as an extension of both these trends (Dillenbourg et al. 2011), due to the increased attention to large groups, parallel teams and the teacher's challenge to manage groups at those various levels.

Stahl et al. (2006) confirm Dillenbourg's (1996) observation that a shift of focus and methods took place: from studying mental models of individual learners cognition to support for (small) groups during collaboration.

2.4.1.2. Group size: Mismatch between theory and practice

An interesting problem in CSCL is that most of the *theoretical knowledge* regarding collaborative learning originally derived from observing small groups or pairs in lab situations (Dillenbourg et al. 1996; Stahl et al. 2006), whereas *in practice* most teaching happens (a) at classroom scale and (b) in the wild. This double mismatch has implications for design:

Firstly, while much is known about how pairs and small groups collaborate by sharing goals and resources, less knowledge is available about the dynamics of sharing in large groups. This makes it difficult for designers to predict how large groups will appropriate a design.

Secondly, given a lack of large-group evidence, how can designers be sure that the predictions made by small-group theories will hold in large group settings? For instance, socio-cognitive conflict theory (Doise et al. 1984) suggests that inducing disagreement in a pair will encourage a productive dialogue that is conducive to learning - as demonstrated, e.g., in Dillenbourg's (2002) ArgueGraph intervention. However, if the same was attempted in a larger group, could we expect the same kind of dialogue to emerge? If so, would it include all participants?

Although studies that systematically compare group size are rare, some qualitative differences in collaboration between dyads, triads and larger groups are undeniable (Stahl 2006). This can easily be illustrated using the above example of disagreement. In a pair disagreement implies that each individual faces exactly one peer with a different view, whereas in a triad this is not guaranteed. Triads enable more complex configuration than pairs, including asymmetry, neutral positions and the possibility of one member disengaging while the other two are arguing. A range of other factors may become relevant in larger groups, such as joint disengagement of subgroups (Kharrufa et al. 2013) or effects of minority dissent on team innovation (De Dreu and West 2001). These and other factors may impact collaborative learning benefits for individuals and/or the whole group.

Given these uncertainties about large-group behaviour, how can instructional designers protect themselves (or rather the learners) from unforeseen, emergent group dynamics? Three major points come to mind:

1. Being aware of well-known large-group social phenomena probably helps.
2. An iterative strategy can be used to identify recurring, detrimental patterns and work out effective solutions.
3. Designs should be sensitive to factors that increase the teacher's capability to manage the group as a whole (Dillenbourg and Jermann 2010).

2.4.2. Collaborative whole-class pedagogies

Among the many pedagogical approaches that focus on peer collaboration, some aim to reconcile small-group collaboration with whole-class awareness. For instance, the SynergyNet multi-tabletop classroom was designed to allow the teacher to switch between both (Higgins et al. 2011).

The “jigsaw classroom” model (Aronson et al. 1978) suggests that (a) the class is divided into small teams, (b) each team is tasked with investigating one of several (complementary) aspects of a topic and (c) finally all the teams reconvene to share and combine their findings - like putting together pieces of a jigsaw puzzle. This pattern has supported various inquiry-based approaches, including outdoors (“treasure hunt”) activities using Augmented Reality simulations (Dunleavy et al. 2009). In contrast to splitting up the group, some notable classroom-based activities, particularly embedded phenomena and some participatory simulations, emphasise engaging the whole classroom simultaneously as one big group.

Integrating the benefits of both small-group collaboration with the benefits of whole-class collaboration (or at least whole-class awareness) has been high on the agenda in CSCL. As recent studies have concluded, integrating a number of small group tasks into a whole-class activity is not straightforward (Dillenbourg et al. 2012). Particularly, if subgroups work in isolation there is a risk of disengagement and lack of whole-class awareness (Kharrufa et al. 2013).

2.4.3. What obstacles impede whole-class collaborative activities?

It is worth asking what circumstances may have prevented collaborative pedagogies from being adopted in schools at large scale. Arguably, many commercial producers of learning materials have little interest in advocating that students share resources. The academic literature suggests a variety of other factors:

1. Orchestrating collaborative activities is challenging for teachers.
 - a) Teachers are forced to divide their attention in order to monitor and assist multiple teams (Dillenbourg and Jermann 2010).
 - b) A whole class can appear chaotic when they all move around (Moher et al. 2005).
 - c) Teachers need more support in the design and implementation of collaborative learning to translate knowledge about collaborative learning into effective practice (De Hei et al. 2014).
2. Technological tools for collaboration are not always effective.

- a) Some tools require too much time (or other resources) to be usable in practice (Dillenbourg et al. 2012).
 - b) Some tools are not designed for the realistic classroom contingencies that require flexible adaptations by teachers (Dillenbourg et al. 2012).
3. Some features of school environments are not amenable to collaboration.
- a) Room layouts (e.g. front-facing versus centred, etc.) impacts the type of interactions among students (Higgins et al. 2005; Mercier et al. 2014). Since these factors cannot be changed easily during a lesson, preference is usually given to a static front-facing configuration that best supports the predominant mode of teaching, i.e., chalk and talk.
 - b) The fragmentation of class time can be a problem. Some collaborative activities (e.g. Moher et al. 2005) take more time than a standard lecture slot, which may potentially require some logistic overhead and coordination with other classes and teachers.
4. Institutional, political and research traditions reinforce the status quo.
- a) Research has been “largely predicated on a traditional view of ‘chalk and talk’ learning in standardised ‘one size fits all’ institutions” (Higgins et al. 2005)
 - b) Traditional content of assessment does not fit well with the higher-order learning that collaboration tends to support (Roschelle et al. 2000).
 - c) Some teachers may need additional training to improve their deep understanding of interactive technologies and adapt their pedagogical style in order to harness what the technology can do (Moss et al. 2007).

2.4.4. Classroom orchestration

Several authors have argued that ‘orchestration’ is a useful metaphor, rather than a definition of a well-delimited research problem (Dillenbourg and Jermann 2010; Prieto et al. 2011). When researchers speak of orchestration in a classroom context, they typically refer to one of the following two concepts:

1. the instructional and/or technological *design* of learning activities;

2. the challenge of *managing* such activities in real time, usually involving one or several human facilitators (teachers, experts, coordinators) to guide and direct a group of people (e.g. students).

This distinction is pointed out in Dillenbourg and Jermann's (2010) influential article on technology for classroom orchestration.

"In music, orchestration refers to writing the score that an orchestra will play. It does not refer to the activity of the conductor when the orchestra is playing. The metaphor is applied in a more orthodox way to computer science where 'service orchestration' [...] refers to the definition of some workflow or architecture that connects different systems (namely web services). Applied to the educational context, the proper meaning of 'orchestrating' would correspond to instructional design and not to the real time management of classroom activities." (Dillenbourg and Jermann 2010)

Depending on the context, the word has been more or less conflated with either technological infrastructure, instructional design or the real-time management of classroom activities. To avoid confusion, this thesis adopts the terminology proposed by Prieto et al. (2011) who conceptualise orchestration as an umbrella term that comprises design as well as real-time management - among other aspects - as explained below.

2.4.4.1. Prieto's "5+3 aspects" framework

In his PhD dissertation Prieto (2012) elaborates a conceptual framework about what classroom orchestration is - in the context of technology enhanced learning and CSCL - and how it should be done. According to Prieto, orchestration comprises the following aspects:

1. design/planning
2. regulation/management
3. adaptation/flexibility/intervention
4. awareness/assessment and
5. the roles of the teacher and other actors.

In addition to these five aspects describing *what orchestration is*, the framework identifies three additional themes, summarising prevalent ideas about *how orchestration should be done*. These were named as follows:

1. pragmatism/practice
2. alignment/synergy
3. models/theories.

Resulting from an exhaustive literature review and validated by an international panel of experts (Prieto et al. 2011), this framework can be considered one of the most exhaustive and up-to-date accounts of the orchestration discourse.

Due to its holistic approach, the framework has much potential to inform the design and empirical evaluation of classroom pedagogies. For instance, Looi & Toh used it to evaluate the design of classroom curricula centred around mobile phones (2014).

2.4.4.2. Dillenbourg & Jermann's teacher-centric framework

In contrast to Prieto's holistic framework, Dillenbourg & Jermann's (2010) influential orchestration framework is teacher-centric, i.e., it focuses specifically on the workload of a teacher during a lesson. The authors stress that while lecturing and managing a large group in real time at multiple "planes" (individual, teamwork and whole class), teachers need to cope flexibly with multiple constraints, including constraints related to the curriculum, assessment, sustainability, space, discipline and security.

The authors furthermore argue that the teacher's role entails far more than what is conveyed by the typical concepts of 'facilitation' and 'management'. Consequently, rather than viewing the teacher as a mere facilitator or a 'guide on the side', they propose the metaphor of an orchestra conductor who coordinates the musicians (i.e. the students) while the orchestra is playing:

Teachers are not on the side, they are the conductors, they are driving the whole activity. They are managing in real time the activities that occur at multiple planes. They share their passion for the content. Their body language conveys their interpretation (speed, intensity, ...) to the musicians."

Within this teacher-centric view of orchestration, 15 important "factors" are identified: leadership, flexibility, control, integration, linearity, continuity, drama, relevance, physicality, awareness, design for all, curriculum relevance, assessment relevance, minimalism and sustainability. This framework has informed several notable studies, including the design and evaluation of whole-class applications involving multiple touch screens and large displays (e.g. Higgins et al. 2012; Kharrufa et al. 2013).

While it is currently uncommon to refer to teachers as “conductors”, the approach described in this dissertation owes much to the idea of designing learning activities such that they can be “conducted” efficiently by a teacher. It stresses the view that design and “conducting” are equal partners in the shared effort of orchestration.

The idea of designing technology to reduce the teacher’s “*overall orchestration load*” is a common connotation of “*design for orchestration*” (Dillenbourg et al. 2011). Conversely, technology’s potential to increase, rather than reduce work for the teacher has also been recognised. To avoid overloading teachers and students with technical demands, recent design explorations have aimed at simplicity. For example, the awareness-supporting “Lantern” device (Fig 2.11) was designed to allow students to call for help simply by pressing the device to change its colour. The colours allow a teacher or teaching assistant to see at a glance who needs help. The design was shown to produce benefits in awareness and time management (Alavi et al. 2009).



Figure 2.11.: The Lantern (Alavi et al. 2009)

2.4.4.3. Orchestration as interoperability

Another emerging aspect of the word orchestration in the CSCL literature is that of creating infrastructures to make technical components work together more smoothly. For instance, improving the interoperability between established CSCL tools and devices, such as Virtual Learning Environments (VLEs), smartphones, whiteboards, etc. (e.g. Prieto 2012; Martinez-Maldonado et al. 2013). Some of this research also considers long-term processes that span multiple lessons and/or multiple classrooms. By contrast, the focus of this dissertation lies on orchestration at the scope of the *individual classroom activity* that is enacted within a single lesson.

Dillenbourg et al. (2011) describe orchestration as akin to usability, whereby “*the classroom is the user*”. This analogy is based on a more general understanding of orchestration as

“the process of managing a whole learning group in such a way as to maintain progress towards the learning outcomes and improvement of practice for all” (e.g. Moon 2001).

As several authors have pointed out (e.g. Dillenbourg and Jermann 2010; Prieto 2012), the word orchestration does not represent a precisely defined research problem but is rather understood as a metaphor used by different authors in different context to mean slightly different things.

“[m]etaphors are not correct or incorrect; the question is whether they are useful or not” (Dillenbourg and Jermann 2010)

A common theme, however, is that the focus of analysis lies on the complex interplay between groups, individuals and external constraints.

2.4.5. Second Outcome: Roles of the teacher and technology

Looking back at the literature of CSCL and simulation-based learning reveals a variety of approaches that assign different roles to the teacher and technology. Despite a diversity of teacher-centric and technology-centric perspectives, some themes are nevertheless widely acknowledged. Concluding from the review so far, it is possible to separate teachers’ “essential” roles from what is “realistically” expected from them in many real settings (Table 2.2).

Eventually it is the teacher’s responsibility to achieve the 15 factors (Dillenbourg and Jermann 2010). The role of technology should be (a) to support, but not overwhelm the teacher in this endeavour, whilst also (b) supporting collaboration among students, by promoting equal information, awareness and control.

These conclusions were adopted in this PhD project as starting points for the empirical studies. While the first empirical study focuses on the “essential” responsibilities, the second study takes the “realistic” responsibilities into account. Both studies aim to refine the above conclusions which, per se, are too abstract to function as practical design recommendations.

| Teacher's role | essential | realistic |
|----------------------------|--|---|
| Before a simulation | The teacher needs to provide appropriate background information (e.g., through lectures or reading assignments) and a specific introduction, explaining the scenario, purpose and rules of the activity. | In some settings it is typically the teacher's responsibility to prepare the activity, including setting up the technology. |
| During a simulation | The teacher acts as the "conductor", rather than a mere "guide on the side" or "sage on the stage". This includes facilitation and may also include ad-hoc lecturing and leading discussion. | Conducting implies juggling many constraints regarding curriculum, assessment, sustainability, space, discipline and security. With complicated technologies, teachers often double as assistants and administrators. |
| After a simulation | The teacher should debrief participants collectively or individually. | Time constraints in large-group settings may favour collective, rather than individual debriefing. The teacher has to take down the activity, including collecting materials, resetting and storing the technology. |

Table 2.2.: Overview of the teacher's role with regard to simulations in real settings.

2.5. Part four: Human-Computer Interaction

Although PCs are still the dominant form of computing device in schools, researchers and industry see much potential in using other forms, such as interactive whiteboards (Moss et al. 2007), multi-touch tabletops (Mercier et al. 2014), tablet computers (Clark and Luckin 2013), smartphones (Looi and Toh 2014), and wearable computers, including optical head-mounted displays, such as Google Glass ².

Given the increasing variety of available form factors and affordances, how can designers decide which types of technology are best suited to whole-class collaboration?

This section reviews a variety of classroom-based approaches from the perspective of multi-user interaction with technology, particularly non-desktop systems. The section concludes with a summary of lessons learned and implications for design, including some caveats, questions and starting points. Particularly, it is concluded that some form of public display is probably needed to create a sufficient level of mutual awareness between all participants and the teacher. Moreover, a combination of ambient displays and shareable touch-screens is considered as a promising starting point, with participants being able to walk around and interact opportunistically with the technology and peers. Questions are raised regarding how to distribute work, information and control among participants, how to encourage equitable participation and how to involve the teacher most effectively.

2.5.1. New technologies

The advent of ubiquitous computing (Weiser 1991) has encouraged system designers to think far beyond the traditional idea of personal computing that assumes a single user per device.

Interactive installations in public spaces, for instance, have challenged the traditionally static notion of *the user*, as groups or individuals can move dynamically from being passive bystanders to being active participants and vice versa. In the context of an interactive display located in a public social setting, Brignull and Rogers (2003) identified three different *activity spaces* between which people moved back and forth: peripheral awareness, focal awareness and direct interaction. Similarly, a public tabletop application for group decision making was found to attract groups in a staggered way, rather than

²www.eduglasses.com

all group members at once - leading to different kinds of interactions than anticipated (Marshall et al. 2011).

Arguably, some of the insights from public settings can be transferred to classroom orchestration. Although classrooms are not really “public” - usually the doors are closed - there are some commonalities in the way that people may need to coordinate their attention between interactive technologies and the activity of other people. Understanding the mechanisms (psychological, social, behavioural, technological, etc.) that can support desirable patterns of interaction is a goal that both UbiComp and classroom orchestration pursue: Whether *the classroom is the user* (Dillenbourg et al. 2011) or *the public is the user* makes little difference when studying the question of how groups share information near an ambient display or device that can be walked up to and used spontaneously.

Ambient displays (Ishii and Ullmer 1997) and spatially fixed, kiosk-like digital hotspots shared by a whole class can enable new forms of classroom teaching. Much of Moher’s work, for instance, hinges on the use of such entry points that are public within the context of the classroom. Those include the shared PocketPC measuring units in RoomQuake (Moher et al. 2005), the monitors in WallCology (Moher et al. 2008) and more recently the RFID-enhanced food patches in Hunger Games (Gnoli et al. 2014). Although these papers did not explicitly focus on mechanisms of shared interaction, such as the ‘honey-pot’ effect (Brignull and Rogers 2003) or over-the-shoulder learning (Twidale 2005), it is nevertheless arguable that such mechanisms were instrumental to the effectiveness of the designs.

2.5.2. The rise of touch-screen technologies in classrooms

Systems like SynergyNet exemplify a research-driven approach that starts with collaborative pedagogy as an ideal goal and then uses the design of shareable technologies and activities in pursuit of this goal. Success has been demonstrated in isolated trials with relatively small groups of students - usually far fewer than 20 (e.g., 16 students in Higgins et al., 2012) - and teachers who were keen to try out new technologies and new ways of teaching.

By contrast, in many real school settings there is a tendency to simply appropriate a new technology to do essentially the same kind of chalk and talk teaching (Rick and Marshall 2010). This has been corroborated by a plethora of studies, including a large-scale evaluation of how interactive whiteboards were used in secondary schools (Moss et al.

2007). The researchers warned that, unless teachers are trained to develop a deep and reflected understanding of the pedagogical affordances of whiteboard technology, it can:

- *“Reinforce a transmission style of whole class teaching in which the contents of the board multiply and go faster (sic), whilst pupils are increasingly reduced to a largely spectator role” and*
- *“Reduce interactivity to what happens at the board, not what happens in the classroom.”*

Indeed, the evaluation showed that the interactive features of the whiteboards were mostly used by the teacher, rather than students. No overall effect of the whiteboards on student performance was found. The initial enthusiasm turned out to be rather short-lived. Evidently, introducing shareable technologies into classrooms in and of itself does not automatically engender more collaborative pedagogies. Against the background of this sobering experience, how should research respond to the current enthusiasm about introducing tablets in classrooms on a large scale? In 2013 the UK Institute of Education published a review on iPads in classrooms, reporting on early adoption scenarios and expectations from researchers, parents, teachers, industry and other stakeholders (Clark and Luckin 2013). Despite being remarkably exhaustive in some aspects, such as the popular idea of replacing textbooks by giving each child an iPad - issues of safety, logistics, finance, etc. - the technology's potential for collaborative use was mentioned only tangentially and in little depth. For instance, while the possibility of passing tablets from one individual to another was generally praised, the studies cited in the review failed to address important question regarding how exactly those superficial technical features were expected to create substantial value in pedagogic practice.

This prevalent lack of attention to collaborative pedagogies in such an exhaustive review is concerning. This is not because that would challenge the validity of the review (which it does not - after all the review's purpose and intention was to summarise what is there, rather than what is considered missing). Rather, it is concerning because it indicates a risk of history repeating itself: Are the tablets in classrooms heading towards the same fate as the whiteboards - to become bright and colourful tools for chalk and talk?

2.5.3. Extending from small-group HCI to large groups

To better understand how mechanisms for collaboration can unfold in the classroom, analysis can start either directly at the classroom scale or study the mechanisms first

in small groups and then consider how they might scale up. User studies have focused on sharing in small groups (e.g. Yuill and Rogers 2012), whereas detailed analyses of large-group sharing at classroom scale are in their infancy.

Few approaches truly capitalise on involving the whole class together as one big group, in the way that Embedded Phenomena and Participatory Simulations do. Many collaborative applications were designed for small groups only, including a plethora of applications for tabletops and single tablets (e.g. Piper et al. 2006; Lyons et al. 2006; Rick 2012a). Using such small-group applications in classrooms is difficult because several instances of the application are needed to allow every student to take part simultaneously. In this case there is a real danger of individual groups being isolated from each other.

2.5.4. Problems with isolated sub-groups

Kharrufa et al. (2013) elaborate on the problem of isolated subgroups and its implications for orchestration. Based on an multi-tabletop classroom evaluation, the authors identified two main problem areas: lack of linearity and lack of teacher's awareness. The concepts are explained below.

2.5.4.1. Lack of linearity

The first problem area identified by Kharrufa et al. revolves around what Dillenbourg and Jermann (2010) refer to as “linearity”, i.e., the goal that *“almost all students can complete at almost the same time maintaining, to a certain level, a synchronized progress through the session”*.

Since some groups finished their task more quickly than others, those groups had nothing to do whilst other groups were still busy with the task. This left the teacher with the challenge of keeping the faster groups occupied in some meaningful way.

2.5.4.2. Lack of awareness

Secondly, since the teacher had to split his or her attention between the groups, each group was “under the radar” while the teacher was further away. This made it easy for groups to disengage from the task (i.e., gossip about what they did last night, etc.) while the teacher was further away and then feign being focused once the teacher returned.

An interesting observation made by the authors was that a gossiping group was very aware of the teacher's location in the classroom, whilst the teacher was considerably unaware of the group. Whether or not the other groups noticed the disengaging behaviour was not discussed, as the analysis focused on the teacher's awareness, rather than awareness between groups. Regardless, it is worth considering what design factors may have caused a situation where students would choose to disengage and were able to do so without the teacher noticing. The teacher being the only person standing may be among the contributing factors, as discussed further below.

2.5.5. Strategies for improving linearity and awareness

To counter these problems, Kharrufa et al. propose a number of design recommendations, mostly speaking from a teacher-centric perspective. The following paragraphs discuss these recommendations and complement them with further considerations from participant-centric perspectives.

2.5.5.1. Supporting linearity

To improve linearity, Kharrufa et al. essentially recommend giving the teacher more flexible control over the system. In summary:

1. The teacher should be able to adjust the task goal or difficulty for each group individually, either during the session or before.
2. The teacher should be able to force the application to proceed to the next stage, even when the predefined conditions have not been met.

In addition to these teacher-centric recommendations, one might also consider participant-centric solutions, e.g., providing opportunities for the faster groups to assist the slower groups at other tables. Such peer support could benefit the assisting students as well as the ones receiving the support.

2.5.5.2. Supporting the teacher's awareness

To increase the teacher's awareness of the group (particularly, the group's progress, quality of work and level of participation), Kharrufa et al. make the following design recommendations:

1. Using a public display and/or a teacher's private device to visualise key progress metrics and outcomes of each table.
2. Visualising each table's history of progress (in addition to the current state) on the tabletop display.
3. Using touchscreen technology that allows the software to identify each user, such as the DiamondTouch (Dietz and Leigh 2001), to encourage equal participation, regulate collaboration and reduce the chances of pretending to work only when teachers are looking.

Much recent work has been concerned with the potential of using visualisation to increase awareness in the classroom. Many of the tools that were evaluated in authentic classroom settings have focused on the teacher's awareness of the students' progress (e.g. Alavi et al. 2009; Martinez-Maldonado et al. 2012).

2.5.5.3. Supporting awareness between peers

Less classroom research has focused on increasing awareness between peers. The hypothesis that mutual awareness can encourage equity of participation has been supported in small-group contexts. For instance, Bachour et al. (2010) evaluated a tabletop system that provides a group with visual feedback on each individual's amount of verbal activity. Such feedback can also be based on other measurements, such as physical activity and touch interaction (Martinez-Maldonado et al. 2013).

However, it is not clear how usable and useful these techniques might be in real classroom situations that are more complex and visually demanding. Conceivably, the additional cognitive load on participants might well outweigh the benefits in such contexts. Besides, there are technical issues with measuring the activity of individual participants. For instance, identifying individuals computationally requires them to either remain in a fixed position - as e.g. with the ReflectTable (Bachour et al. 2010) or the DiamondTouch - or wear special mobile devices, neither of which is generally practical in classrooms.

There are, however, indicators that face-to-face awareness (and thus equity of participation) can be supported in less explicit ways. Some whole-class approaches (e.g. Colella 2000; Moher et al. 2005) encourage students to move around in the room. This potentially creates a situation in which encounters between participants can spontaneously emerge anytime anywhere in the classroom. Although studies have not investigated this issue explicitly, one may hypothesise that the kind of under-the-radar disengagement described

above may be less likely to occur in such standing up situations, since participants are never quite under the radar - at least they may reckon that their peers are likely to notice. Besides, it is arguable that standing up situations create fewer incentives to disengage, compared to sitting down, since they imply an option for participants to simply go somewhere more interesting when they feel bored.

It may therefore be hypothesised whether small-group configurations around individual tables and/or digital artefacts are generally more prone to under-the-radar disengagement than whole-class activities. No direct comparison has been made so far between those two approaches in a multi-device classroom context.

2.5.6. Third outcome: Starting points for interaction design

The third part of the literature review suggests a number of tentative recommendations for interaction design. These are:

1. To support fluid collaboration in the whole group, the technology should be designed to carefully constrain information, control and awareness.
2. Shared touch surfaces (tablet size or larger) and ambient displays are probably the most appropriate technologies for supporting a general variety of whole-classroom problem solving tasks.
3. Participants should be allowed to interact with each other opportunistically, supporting a community of practice.
4. Team-based group structures can support collaboration as well as competition. Both can be beneficial factors for engaging participants in activity and dialogue.
5. Linearity and cross-plane integration (Dillenbourg and Jermann 2010) should be supported by the technology.

2.6. Summary and conclusion

This chapter reviewed relevant work in the areas of simulation-based learning, CSCL and HCI in order to inform a new design approach for supporting whole-classroom collaboration. The results of the review can be summarised in three main outcomes (Table 2.3):

| Outcome | Description |
|----------------|--|
| First Outcome | A pedagogical approach that aims to support participants' higher-level learning, particularly decision-making, debate and reflection. The approach capitalises on shared, interactive computer simulations embedded in a context of lectures and whole-class discussion. |
| Second Outcome | Principles for orchestration, involving specific roles for the teacher and technology. |
| Third Outcome | Recommendations for how interaction design can support fluid participation and orchestration. |

Table 2.3.: The three main outcomes of this literature review

Further questions remain to which the literature review provided only starting points but no definitive answers. Drawing these together leads us to the four research questions Q1-Q4 (see Chapter 1) that needed to be answered through empirical studies.

Chapter 3.

Methodology

This chapter introduces the overall methodological approach and explains why certain research and design methods were preferred over others. In particular, an iterative approach is proposed, with a clear focus on pedagogic needs and teachers being part of the design team. Furthermore, the chapter explains the methods of data collection and analysis that were used in the studies.

3.1. Design-based research approach

The overall research aims suggest a design-based research approach as is common in HCI and the learning sciences. Based on the conclusions drawn from the literature review, it was decided to design, develop and evaluate a number of whole-classroom applications in the wild to devise a generalisable framework of design factors that support the learning goals. The requirements based on the literature review can be summarised as follows:

- Give the group an important problem to solve;
- Establish the necessary low-level knowledge (via lectures) that the group needs in order to make meaningful decisions;
- Provide a technical infrastructure that encourages sharing and collaboration and provides feedback on decisions;
- Design for orchestration to allow the teacher to stay in control.

3.1.1. From concrete design to abstract concepts and back

As illustrated in Figure 3.1, the overall design-based research strategy can be conceptualised as three consecutive processes, here termed instantiation, abstraction and validation:

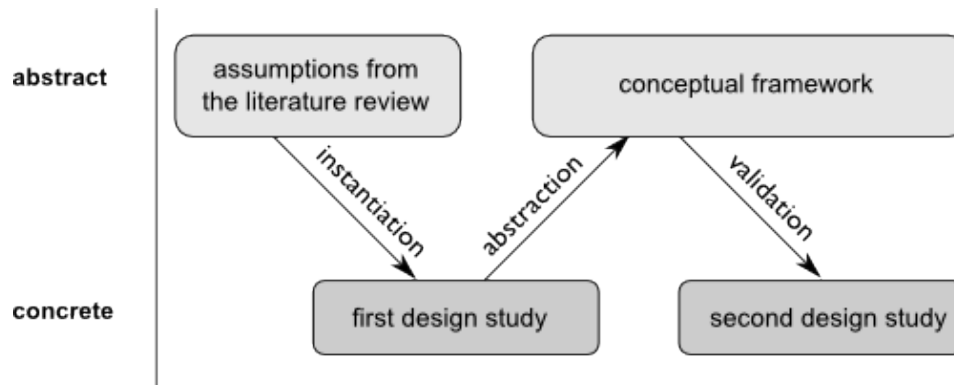


Figure 3.1.: Overview of the approach taken in this research

1. **Instantiation:** First, a concrete learning activity was designed and instantiated in a real setting. Group interactions and participant feedback were analysed using qualitative methods.
2. **Abstraction:** By conceptualising these data in abstract terms, a first step was made towards making meaningful generalisations. This analysis involved some interpretive judgements, including:
 - a) What learning goals are and feasible?
 - b) What system features and human factors are conducive to these goals?
 - c) How should the orchestration load be balanced between the teacher and the technology?
 - d) What design factors are likely to “work well”?
3. **Validation:** In order to challenge and refine these initial judgements - particularly the last point - a second design study was conducted in a different setting, using similar learning goals and a similar set of technologies. Among other goals, such as finding ways to improve the sustainability of the technical framework, the main goal of this validation study was to investigate the following questions:
 - a) Will the conclusions from study 1 hold true in a different context with a different student population?

- b) Are the initial framework's concepts and terminology useful in a different design case, particularly with teachers who were not involved in the initial design?
- c) Will the framework's recommendations lead to a successful design?

Findings from both studies were used to iteratively refine the conceptual framework and the overall conclusions.

3.1.2. Design layers: The “Triple-i” Hierarchy

Taking to heart Dillenbourg’s notion that “the classroom is the user” (Dillenbourg et al., 2011), the ultimate goal of classroom design should be to enable and facilitate a workflow of learning and teaching that suits the classroom as a whole. Aspects of group interaction and technology are considered subsidiary, albeit instrumental, to this ultimate goal. One can broadly summarise this as a hierarchy of three layers: instruction, interaction and interfaces (see Figure 3.2).

1. **instruction** (aka instructional design / pedagogy) concerns the teaching and learning needs. It includes specifying the learning goals, the structure of a lesson and how to assess learning.
2. **interaction** (aka human-human interaction) is primarily concerned with mechanisms for collaboration (Yuill and Rogers 2012) and other social mechanisms between students and interactions with the teacher. Relevant aspects may include team play, competition and joint cognitive tasks, such as problem solving, searching, discussing, hypothesising etc., that are considered beneficial to the learning needs.
3. **interface** (aka human-computer interaction) focuses on usability on a person-to-device level. This layer involves choices of hardware (i.e. types and number of devices for students and the teacher), the design of input and output elements, on-screen information and control. Enabling multiple users to share a device or display is also a key concern in this layer.

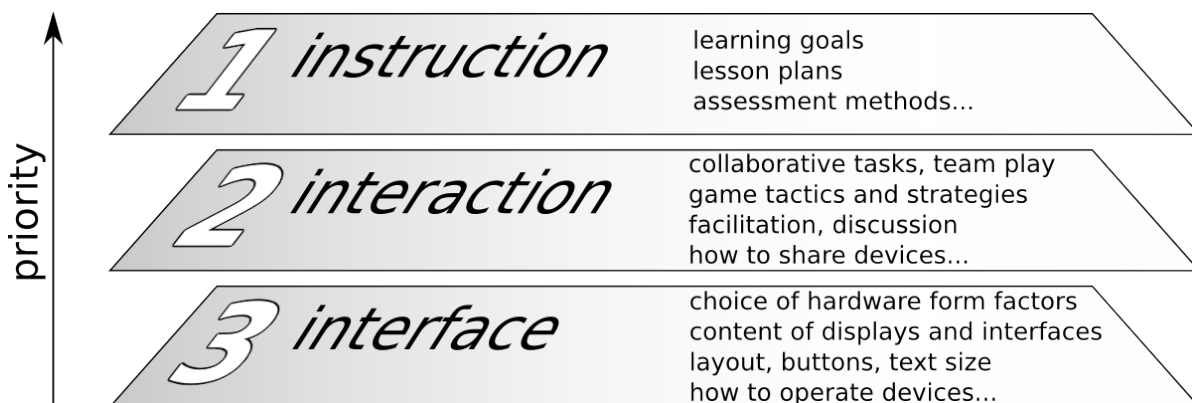


Figure 3.2.: The “Triple-i” hierarchy of design goals

This hierarchy emerged during the early stages of design. Its purpose is twofold:

- **to guide design:** The hierarchy suggests a certain order in which to prioritise design goals. Specifying the instructional needs (layer 1) is top priority, since the purpose of all further design decisions should be to fulfil those needs. Similarly, design choices regarding interfaces (layer 3) are meaningful only to the extent that they serve a defined purpose in a shared context of group interaction (layer 2). Overall, the lower layers are thought to serve the higher layers. This order ultimately defines how design decisions should be made - although inspiration may sometimes also go the other way round.
- **to guide evaluation:** The same priorities can be applied to the evaluation of sketches and prototypes, by asking (1) whether the design goals in each layer are relevant and well-defined and (2) whether each of the lower layers effectively supports the one above it.

A likely implication of adopting this method is that over the course of an iterative design process, design aspects related to the higher layers will mature before the lower ones.

The following chapters will frequently make reference to the Triple-i hierarchy, using the abbreviations i1, i2 and i3 for the three layers.

3.1.3. Teachers taking part in the design

Teachers who took part in the studies were prepared to facilitate the activity multiple times, rather than one-off. It was conjectured that the prospect of sustained use would encourage the teachers to apply a realistic sense of judgement when contributing to design decisions.

The total effort a teacher may invest (preparation work, time to provide feedback, etc) is limited. If we design for heroes, we lose scalability (there are few heroes) and sustainability (heroes get tired). (Dillenbourg 2013)

Along these lines, the approach aimed to create sustainable products that the teachers themselves (or their colleagues) would be ready to use on a regular basis.

Planning to conduct an activity multiple times as part of their work implied a considerable amount of personal and professional commitment from the participating teachers. It was hypothesised that being personally and professionally invested in the potential cost and benefit of the design outcome would incentivise teachers to bring their true needs

to the design table. As equitable members of the design team, they were encouraged to highlight any requirements in support of their long-term needs. Conversely, they were encouraged to detect and oppose any design factors that were somewhat in conflict with their long term needs.

3.1.4. Iterative design and prototyping

When designing novel applications, it is often desirable to assume an iterative strategy. Such an approach emphasises repeated cycles of planning, prototyping and testing, as opposed to implementing the entire application based on initial requirements.

The philosophy behind this approach *rejects the idea that every relevant aspect of a design can be predicted in advance*. Rather, it stresses the belief that some user needs are unpredictable and that users may appropriate the design in unpredicted way. For instance, Marshall et al. (2011) report on a multi-user tabletop application that was deployed in a public setting. Based on prior ethnographical observations in the setting, the designers had assumed that visitors would approach the tabletop as coherent groups. However, when the system was actually deployed, it turned out that visitors often approached the system one by one or in staggered groups. This defied the system's design as a multi-user application and led to considerable confusion and many instances of unsuccessful use. Findings like this emphasise the importance of early and continuous testing in real settings in order to continuously revalidate the design assumptions.

To ensure that the design goals were eventually achieved within the constraints and unpredictabilities in the wild, both studies adopted an iterative approach, involving multiple iterations of design, evaluation and redesign. Numerous sketches and prototypes were tested in the lab for their robustness, usability, shareability, and pedagogical effectiveness.

3.1.4.1. Generating knowledge from design

From these considerations followed a research approach that considered user tests as the main (although not the only) source of academically significant knowledge. By conducting rigorous tests with representative users under representative conditions, designers and researchers can derive new understandings. While some of the findings may be specific to the *concrete* use scenario at hand, others may bear general significance to an entire,

abstract class of potential use scenarios, i.e., to future design and thus the academic community.

Taking both the concrete and abstract perspective into account during the process is a methodological principle in this research. The rationale is that neglecting the abstract perspective would result in a lack of significant knowledge to report to the community, whilst neglecting the concrete perspective would likely result in a dysfunctional product that may not be worth reporting.

3.1.4.2. Continuous designing until goals are met

In order to be able to report on a functional product, both design studies were guided by the intention to keep designing and evaluating until the design goals were satisfied. Obviously, this approach bears a certain risk that the design might go on forever if the goals cannot be met. Alternatively, if one kept delivering dysfunctional products then educational settings might cease to offer opportunities for evaluation in the wild. Therefore, this approach requires persistence from the designer/researcher and the ability to go back and change direction if an idea is heading nowhere.

3.1.4.3. Dealing with uncertainty

Judging the significance (specific or general) of any particular finding will always depend on interpretation in context. Based on these interpretations, previous design assumptions may have to be revisited and the design may take a turn. What follows from this circumstance is that some important aspects of the design are essentially unknowable from the beginning:

- the final shape of the application
- the number and kind of changes that the design undergoes over time
- the relevant variables that the design goals depend on

In this sense, the iterative design approach differs markedly from the hypothesis-driven experimental approach that aims to generate a pre-specified type of outcome, based on the assumption that all but a few variables are knowable (and in fact known) in advance.

3.2. Analytical focus

The paramount goal of this research was to make collaboration more feasible at classroom scale. While the design aimed to address this problem of feasibility by proposing solutions, the primary goal of analysis was to test these solutions regarding their actual feasibility.

This included (a) evaluating whether the proposed design goals were indeed achieved and (b) trying to surface any unexpected factors that were relevant to achieving the design goals. A variety of factors had to be considered in this regard, including practical aspects of orchestration, technical robustness and usability of the prototypes, participant behaviour, facilitation strategies and perceived individual learning.

3.2.1. Engaging, enjoyable, usable and useful

It is widely accepted that active engagement with a topic has positive effects on learning, as the literature review has shown. Therefore, this research is not concerned with the question whether or not engagement is pedagogically effective, but rather how engagement can be enabled and encouraged. This question directs the focus of analysis to aspects of human interaction, rather than pre and post test (Stahl 2006).

Unfortunately, the word “engageability” lacks aesthetic grace. Otherwise, it would be well suited for describing the kind of use quality that this research aims to support and measure. Especially Chapter 4 is about the question “how can people engage”.

Achieving joint engagement in a large group plus teacher is not trivial, as the literature on classroom orchestration has shown. Identifying factors that benefit or hinder relevant forms of engagement in this context is considered a significant contribution to knowledge.

That said, engagement per se is not sufficient to warrant successful learning. Firstly, one must differentiate between relevant and irrelevant engagement. For instance, in a tablet-based learning activity about climate change, participants getting excited about the technical features of tablets would count as *irrelevant* with respect to the learning goals. Rather, it is desired for participants to engage in meaningful, critical dialogue about the topic on multiple knowledge levels, such as facts, concepts, processes and metacognition (Kratwohl 2002).

Secondly, it makes sense for evaluation to consider aspects of engagement in the context of other measures, such as usability, enjoyability or perceived usefulness. The latter may be judged differently by learners themselves, teachers and other stakeholders, such

as course designers. The combined explanatory power of these measures is considered sufficient for answering the research questions.

3.2.2. Relevant metrics

To operationalise engageability in the evaluation of a particular design, the following set of questions can be used as a starting point:

- Does the activity include every student?
- Does the activity encourage all students to contribute equitably?
- Does the activity keep students focused on the learning goals, or is it partly distracting or boring?
- Are students' discussions meaningful and productive?
- Is all dialogue during the activity focused on the topic - as opposed to, e.g., students chatting about unrelated topics (see, e.g., Kharrufa et al., 2013)?
- Does the activity allow the teacher to monitor students' participation and quality of contribution?
- Do students perceive their participation in the activity as useful for learning?

Qualitative methods were used to better understand the social factors, design factors and teaching practices that are relevant to these aspects. This approach differs notably from using pre and post tests or control groups, which were not used in my work.

3.2.3. Why not use pre and post tests?

Pre and post tests aim to objectively measure individual students' learning gains, by testing the individual before and after an intervention. Such tests usually focus on content-specific comprehension, retention and skills (e.g., Moher, 2005), rather than higher level learning.

Although pre and post tests are commonly used in other educational contexts, this method was not considered appropriate for assessing the type of learning and design objectives pursued in this work. A number of theoretical and practical reasons are elaborated below.

The first set of reasons are of a theoretical nature:

1. Some CSCL researchers argue that “[...] *one cannot measure the learning - even the individual learning - that takes place in collaborative situations with the use of pre- and post-tests that measure capabilities of the individuals when they are working alone.*” (Stahl et al., 2006, p.8). The argument is based on Vygotsky’s (1978) concept of “zone of proximal development” which suggests that individuals have different mental capabilities in collaborative situations than when they are working alone.
2. Pre and post tests fail to provide insight into relevant details of a session, such as drivers of engagement, equity of participation, peer interactions and the influence of the teacher.
3. Pre and post tests measure lower-level (content) knowledge and skills, whereas the focus of this PhD project is on higher-level learning.
4. The particular designs focused on some complex, ill-defined problems that did not have any objective “right or wrong” answers (particularly, global and personal spending decisions). This does not conflict with higher-level learning in any way - but it is hard to reconcile with the kind of well-delineated and objective metrics that pre and post tests require in order to be useful.

Furthermore, besides the above theoretical considerations, the practical feasibility of pre and post tests in this context was limited by the following factors:

1. Since the activities were designed as a combination of teaching methods (including simulation, lecture and discussion in rapid succession), any measured learning gains could only be ascribed to the the whole mix - providing little insight as to the contribution of each part, e.g., the simulation.
2. Effects may not be visible considering the brevity of the designed activities. Studies that found significant effects of classroom collaboration on learning (relative to traditional teaching methods) have typically focused on sustained exposure over a prolonged period (Nussbaum 2008). Since our design studies aimed for rather short activities, many such activities would have to be developed, which was not feasible within the scope of this PhD project.
3. For the same reason, measuring long-term skills development, such as collaboration and discussion skills, was not considered feasible.

4. Measuring long-term behaviour change was not considered feasible, either. For instance, it is simply not practical to follow senior managers back into their organisations looking for evidence of individual or organisational change of behaviour or policy.

3.2.4. Why not use control groups?

A common approach for evaluating learning interventions is to conduct controlled experiments with two comparable groups of learners, whereby one group is exposed to the intervention while the other group spends an equal amount of time on a standard or neutral task. Several simulation-based teaching tools have been evaluated this way (e.g., Keller et al., 2006; Wayne et al., 2008).

However, this method was considered practically infeasible in the context of this research, due to issues of measurability (see Section 3.2.3 on pre and post tests) and poor availability of comparable groups in the wild.

3.3. Data collection

Each study was conducted with real users in real settings. To assess whether the systems “worked well” in every relevant respect, different perspectives had to be considered. These include each individual participant’s experience, as well as the large group as a whole and the perspective of teachers and potentially other stakeholders, such as course organisers. This multitude of perspectives is reflected in the different types of data collected.

3.3.1. Types of data

To account for those multiple perspectives, it was necessary to capture and triangulate between different types of data:

- Field notes
- Teacher interviews
- Student questionnaires
- Video and audio recordings
- System logs and screenshots
- Transcripts of debriefing and vignettes

3.3.1.1. Field notes

Notes were taken on paper during each of the evaluations. The notes served various purposes, ranging from technical reminders to detailed ethnographical observations of participant groups.

3.3.1.2. Teacher interviews

Semi-structured interviews were conducted with the teacher after each lab test or in situ session to elicit salient aspects of the teacher’s experience. Subject to available time, another interview was conducted a few weeks later to discuss long-term impressions and reflections in a wider context. The interviews mostly happened face to face and occasionally by phone or video conference.

In the first study (as the teacher was also a researcher) video excerpts and debriefing transcripts were occasionally used to focus on various details of a session.

3.3.1.3. Student questionnaires

Paper-based questionnaires with open questions and Likert scales were used after the sessions and sometimes during the sessions. They had to be short due to time constraints in situ. Therefore, to get the most insight of the questionnaires, redundancy was avoided by (a) avoiding questions that could be fully answered using other methods, e.g., video analysis, and (b) continuously updating and refining the questionnaires from session to session as the research progressed. Nevertheless, some basic questions, such as, “did you find the activity useful for learning”, were important to ask every time and were thus kept the same across sessions.

In addition to the questionnaires on-site, an attempt was made to elicit participants’ long-term impressions through a delayed online survey. However, this method failed to provide a useful amount of data, as only very few participants responded to the voluntary survey.

3.3.1.4. Video and audio recordings

The first series of in-the-wild evaluations capitalised on video analysis of participants’ interactions. Four wide-angle cameras were used for this purpose, one in each corner of the room, to cover the whole room and capture every move of every individual. Three stereo channels were recorded to capture the discussion and the teacher’s interaction with the group. This high resolution data was considered necessary for constructing a detailed and exhaustive account of how groups shared the devices and displays in the physical space.

Later evaluations, including the second study in school classrooms, were more teacher centric and thus focused on field notes and interviews, rather than detailed analysis of participant behaviour.

3.3.1.5. System logs and screenshots

Implementing the prototypes as server-based applications made it easy to track the state of the simulation for later analysis. This included relevant events in the simulations and screenshots of displays at relevant moments.

3.3.1.6. Transcripts of debriefing and vignettes

The debriefing discussions were fully transcribed in the 4Decades study. Moreover, partial transcripts were made of important moments during the activities for detailed analysis.

3.3.2. Amount of data

While both studies mostly relied on the same *types* of data, they differed in regard to the *amount* of data collected. Large amounts of fine-grained data were required in the first study in order to build the conceptual framework and relevant assumptions. Testing this framework and assumptions in the second study required less data.

For instance, at the beginning of the first study there was little knowledge about how large groups could share multiple tablets and displays in the wild. To ensure insight into every participant's movements, eye-gaze and gestures, four video cameras were used and questionnaires covered many details regarding usability. Once the analysis had revealed a useful set of generic patterns, field notes and shorter questionnaires were deemed sufficient for the second study. Moreover, this lean approach to data collection was more suitable to the highly time-constrained setting of the second study.

3.3.3. Lab tests versus in-situ evaluation

The whole research approach was situated “in the wild”, in the sense that both studies targeted real educational needs in real, typical settings. Real teachers working in these settings played central roles in the design and evaluation of the learning activities developed in the course of this research. It was assumed that the participation of teachers would provide realistic insights into the needs and abilities of the given student populations, the constraints of the given settings and thus the potential of design to improve learning in these settings.

A hybrid approach was taken that combined elements of lab studies with in-situ evaluation. Both approaches have different strengths and limitations in the context of multi-user technologies (Rogers, et al., 2013).

3.3.3.1. Aims of the lab tests

While the specific test conditions and objectives naturally varied over time, depending on what aspect of the design was being tested, the following goals were generic to all test sessions:

1. to test whether the proposed design goals were met;
2. to test the prototypes for technical robustness and usability under multi-user conditions;
3. to refine the pedagogy, i.e., giving the teacher some opportunities to practice and try out different approaches to introduction, facilitation and debriefing;
4. to refine the methods of evaluation, e.g. finding suitable camera angles, refining the questionnaires to ensure they were clear, salient and concise.

To source a diversity of critical perspectives on the activity and the evaluation procedures, I aimed to recruit many volunteers from diverse areas of interest, including computing, science, gaming, etc., including fellow researchers, friends and others, mostly sourced from our university. In a friendly lab situation these volunteers were encouraged to raise all kinds of issues that appeared important, including issues that the ‘real’ participants would be unlikely to raise, due to time pressure and social constraints. Moreover, it was expected that some of the lab volunteers would be available for in-depth discussion after the sessions.

3.3.3.2. “Wild” conditions in the lab

The approach here was to test initial designs in lab sessions that were conducted as “rehearsals for the wild”, rather than controlled lab experiments. The goal was to test the prototypes and rehearse the orchestration under conditions that replicate the complexities in the wild.

This involved aiming to recruit large groups of a representative demographic, priming them with a short introductory lecture on climate change and proceeding through the

activity at a realistic pace as if it were in situ. However, the possibilities of approximating real conditions were limited. For instance, giving participants two full days of climate lectures before a lab session was unrealistic. Moreover, some emergent aspects were impossible to simulate, including, e.g., the social dynamics between representatives from rivalling organisations.

3.3.3.3. “Lab” conditions in the wild

In the first study, real seminar rooms were fitted with high-fidelity recording equipment, using multiple cameras and microphones, as if to record a live concert. This was practical and desirable only in the first study, whereas in the school classrooms there was neither the need nor the time.

3.3.4. Choice of settings: From professionals to schools

The idea of transferring insights from corporate learning into classroom contexts is not entirely new. For instance, the concept of ‘community of practice’ (Lave & Wenger, 1990) derives from workplace learning and has inspired a range of innovative pedagogies aimed at classrooms (e.g. Brown & Campione, 1994; Moher, 2005; Peters & Slotta, 2010). Compared to the majority of these approaches, this PhD research project stands out by literally conducting the same type of design study in both settings, using the same pedagogical framework. The analytical perspective gained through this direct approach is rather unique.

Practical reasons also played a role, such as the opportunity to evaluate an activity multiple times in a row with comparable groups and the same teacher. This was possible in the chosen professional training setting but is usually difficult to do in school settings, due to curriculum constraints and scheduling.

While the chosen professional setting offered certain benefits on the one hand, it also entailed some specific constraints on the other hand. The following two subsections elaborate on some of the key benefits and constraints and their implications to the research methodology.

3.3.4.1. Benefits of the professional setting

1. **Scheduling and permission:** Since workshops with new cohorts took place every few months, it was possible to conduct multiple evaluations with comparable groups in short intervals. Being able to test frequently with real users is important when doing iterative design. This tends to be difficult in schools, due to long-term curriculum constraints and the lengthy procedures of getting permission for data collection. In contrast to many previous classroom studies, doing a single one-off evaluation to report on the strengths and weaknesses of the approach was not considered sufficient for this purpose. Rather, the aim was to keep designing and evaluating until the design could be considered mature in terms of its goals.
2. **Practical flexibility:** With the evaluations taking place at conference centres and multi-purpose venues, there was much flexibility for conducting the evaluations. For instance, the researchers were given plenty of preparation time to set up a suitable room with enough cameras and microphones to capture data at high fidelity. Technical staff was constantly available to assist with any problems. Moreover, the venues were free from distractions such as noise from parallel classrooms, which was beneficial to efficient orchestration and data collection.
3. **Participants' communication skills:** Getting clear, explicit feedback from participants is easier when they have strong verbal communication skills. Most of the participants in the first study were senior managers, engineers or academics who had little difficulty articulating their learning experience clearly in full sentences during debriefing discussions and in questionnaires. Their comments could generally be assumed to be honest, serious, critical and to the point.
4. **Participants' polite behaviour** Due to participants' general maturity and experience in working productively in mixed groups (age, gender, culture, etc.), polite social behaviour could be expected during the sessions, allowing the orchestration to focus on learning objectives.
5. **Participants' competence in high-level learning:** In contrast to school settings that are traditionally oriented towards (low-level) recall and comprehension tests, it is common for many professional training courses to focus on higher level learning. Since the latter is more difficult to assess systematically and many professional training courses are not graded, much of the responsibility to assess learning gains tends to lie with the adult participants themselves. Most likely, their judgements about what they learned will be grounded in comprehensive experience with social

learning, include prior professional training, higher education and learning on the job. Therefore, it was assumed that the majority of ratings and comments made by participants in questionnaires, interviews and discussions could be treated as fairly reliable indicators of an intervention's pedagogical effectiveness.

3.3.4.2. Downsides of the professional setting

1. **Pressure to meet high standards:** Professional training workshops tend to be prestigious and expensive. It is therefore normal for participants to expect a top quality learning experience, i.e., one that results in high perceived learning gains, while using participants' time efficiently. A learning activity that falls short of these expectations has low chances of being deployed again. Therefore, the ability to do iterative design through repeated deployment is limited by the need to have a fairly effective prototype to begin with.
2. **Using participants' time efficiently:** Professional training seminars are often condensed time spans during which participants are excused from their jobs, at the expense of their organisations' productivity. Design and research methods need to take into account that participants have little patience for excessive waiting, turn taking or time-consuming procedures of data collection. Moreover, there is little tolerance for technical delays or boring delivery. This boils down to the importance of providing a technically robust, well-rehearsed activity.
3. **Participants' social and organisational bias:** Professionals are often tied within organisational relations, which may sometimes constrain what they say in the presence of other stakeholders, such as colleagues, superiors, competitors, potential clients, business partners, etc.. For instance, it is conceivable that the necessity to appear smart in front of others might prevent some individuals from openly admitting certain difficulties in using an interface or understanding parts of an activity. Other types of bias might include tendencies to over-praise innovative interventions merely to avoid sounding conservative. Although some of these social factors also exist between peers in classrooms, they may be exacerbated in professional settings. Researchers need to be sensitive to these potential biases when evaluating what people say openly (e.g., during a simulation or debriefing) and in confidence (e.g., in questionnaires).

3.4. Qualitative analysis

Both studies used qualitative methods to develop conceptual knowledge and design recommendations.

3.4.1. Overview

The analysis relied on data collected in the lab and in situ with real participant groups. Whereas the first study capitalised on video data, the second study relied mostly on teacher interviews, field notes and questionnaires. Despite these differences in emphasis, both studies applied essentially the same logic of analysis, which is outlined in Figure 3.3.

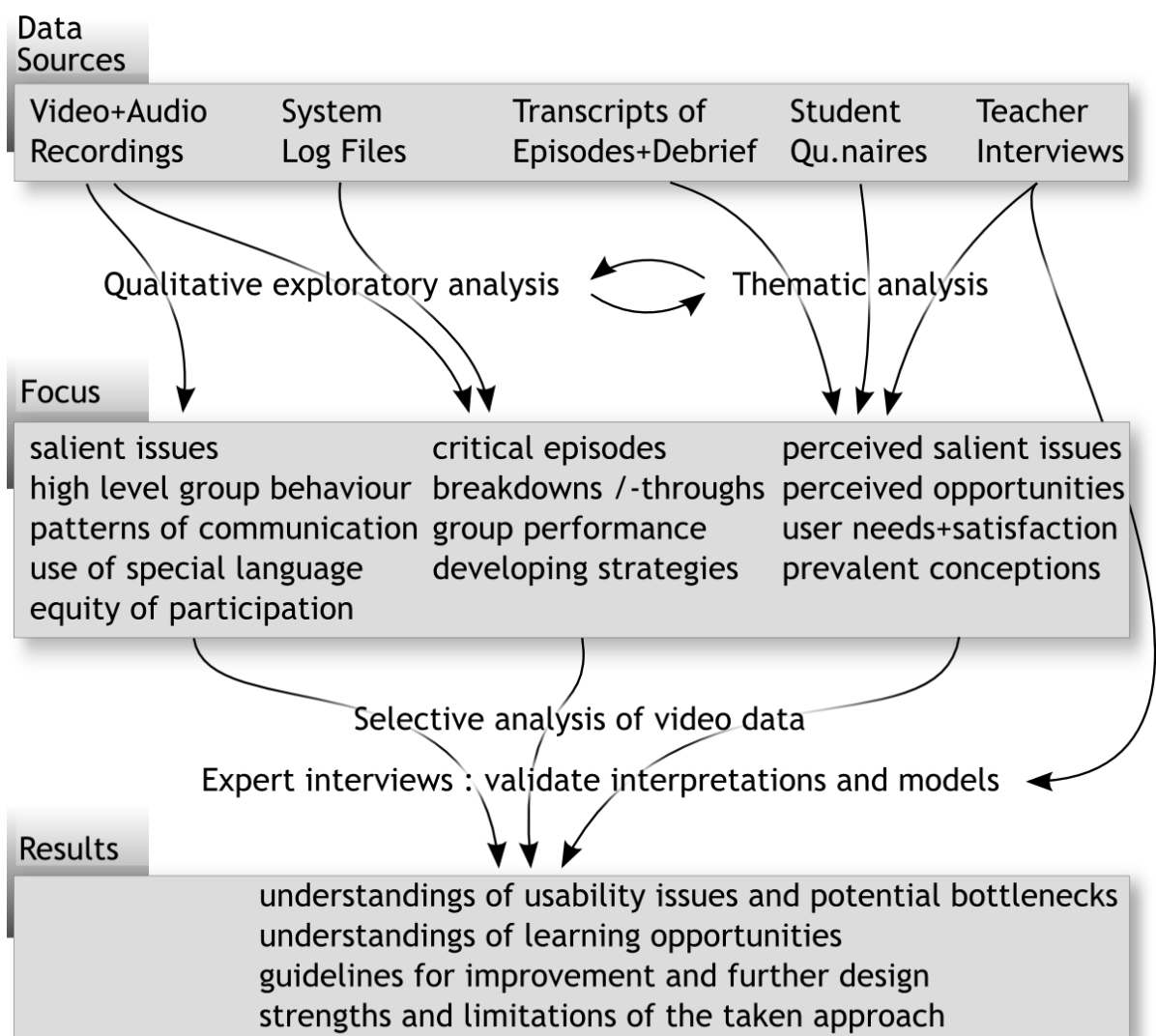


Figure 3.3.: Overview of the qualitative analysis strategy

3.4.1.1. Video analysis

In the first study, which capitalised on video analysis of participant behaviour, video data had to be analysed in several passes at various levels of granularity. This was necessary in order to deal with the sheer density and complexity of interaction, including up to 30 people and a teacher.

The goals were (a) to verify some of the predefined design goals, such as easy usability and equal physical access to peers and technology, and (b) to identify unexpected interactions or mechanisms that were relevant to the learning experience.

For this purpose, spatial configurations of people, physical objects and technology were considered, as well as verbal interactions, gestures and visual awareness. This part of the analysis resulted in (a) a series of macro-level visualisations of relevant behaviours over time and (b) a set of vignettes highlighting salient moments, such as break-throughs and breakdowns in team communication, that notably affected the experience for all or specific participant groups. Some of these vignettes were discussed with other researchers, including the facilitator of the activity. The results are described in Chapter 4.

The second study used detailed video analysis only with regard to the teacher's actions, focusing on how the sessions were orchestrated. Relevant themes of the teacher's behaviour were visualised on a timeline. The third study capitalised involved video only indirectly, insofar as another researcher (Nicola Yuill at ChaTLab, University of Sussex) conducted video-based evaluations in the lab and a school setting. My analysis took into account some preliminary conclusions from those findings.

3.4.1.2. Teacher interviews

Since the teachers were also involved in the design, evaluation of the interviews was strongly tied with the process of developing, testing and documenting the sketches and prototypes.

3.4.2. Scope

Due to the high effort involved in conducting design and evaluation in the wild, the scope of this research was limited to a small number of in-situ sessions. Moreover, some diversity between the available groups and settings was practically inevitable and had to

be taken into account. Therefore, the analysis prioritised detail over generalisation, first aiming at a rich and differentiated understanding of issues within the context of each enactment. comparisons and generalisations between sessions were made tentatively and only in regard to specific aspects where the differences between groups and settings could be neglected. Consequently, no comparative statistical methods were used. Rather, the quantitative parts of the data (notably, Likert scales and coded video data) were mainly used for description and triangulation with other data.

3.5. Summary

This chapter introduced the overall methodological approach and explained why certain research and design methods were preferred over others. In particular, iterative design and evaluation was considered the most adequate strategy, with teachers on the team to ensure that the focus remains on addressing concrete, pedagogic needs.

Figure 3.4 draws out how the following chapters relate to the main research questions, types of data, analysis methods and conducted studies, respectively.

| | chapter 4 | chapters 5-7 | chapter 8 | chapter 9 |
|-------------------------|---|---|---|---|
| Question | Q1 How can the whole group engage with the topic? a) What types of engagement goals are desirable and feasible? b) What mechanisms can effectively support these goals? | Q2 How should one orchestrate the activity? a) What constitutes the orchestration load? b) How should the orchestration load be balanced between teacher and technology to achieve feasibility? | Q3 What are useful abstractions to aid future design, orchestration and evaluation? a) How can we generically describe the factors that make or break the design? b) What are practical recommendations for teachers and designers? | Q4 Are these abstractions also useful when designing for sustainability? a) Is the CLOVER framework applicable in different settings? b) What design factors are critical for achieving sustainability in the wild? |
| Study | 4Decades (lab) | 4Decades in the wild | | UniPad Comfy Birds |
| Data | lab video, field notes system logs teacher interviews student questionnaires debriefing transcripts | in-situ video, field notes system logs, screenshots teacher interviews student questionnaires debriefing transcripts | findings from 4Decades and literature | notes from design meetings with teachers teacher interviews field notes system logs |
| Analysis methods | lab-based prototyping qualitative analysis | detailed video analysis qualitative analysis of system screenshots | reflection on design development of a conceptual framework | lab based prototyping preliminary feasibility evaluations in the wild qualitative analysis |

Figure 3.4.: Overview of questions, data types and analysis methods

In addition to this chapter, some chapters have dedicated “method” sections that focus on specific aspects that are relevant in the context of the particular chapter. Finally, the discussion chapter reflects on some methodological decisions and discuss limitations and lessons learned.

Chapter 4.

Designing a participatory simulation on climate change

4.1. Overview

This chapter covers the initial design and evaluation of the 4Decades simulation. The 4Decades study involved several iterations, prototyping, lab tests and in-depth evaluation in the wild. To provide an easily readable structure, the study is split up into multiple chapters. This chapter focuses exclusively on the lab-based prototyping, i.e., up to the point before the first deployment. Chapter 5 then reports on the evaluation in the wild. Findings from this evaluation are used in Chapter 8 to derive a conceptual framework that is validated in Chapter 9. An overview of these chapters is presented in Figure 4.1, summarising the main research questions (paraphrased), data sources and analysis methods used.

The 4Decades simulation helps groups of policy makers discuss future climate scenarios, using ambient information and shared tablets. The study was conducted in the context of executive training courses that were hosted by the University of Cambridge (Figure 4.2). The activity was facilitated by Dr Stephen Peake who also designed the mathematical model that underpinned the simulation.

| | chapter 4 | chapters 5-7 | chapter 8 | chapter 9 |
|------------------|--|--|--|--|
| Question | <p>Q1</p> <p>How can the whole group engage with the topic?</p> <p>a) What types of engagement goals are desirable and feasible?</p> <p>b) What mechanisms can effectively support these goals?</p> | <p>Q2</p> <p>How should one orchestrate the activity?</p> <p>a) What constitutes the orchestration load?</p> <p>b) How should the orchestration load be balanced between teacher and technology to achieve feasibility?</p> | <p>Q3</p> <p>What are useful abstractions to aid future design, orchestration and evaluation?</p> <p>a) How can we generically describe the factors that make or break the design?</p> <p>b) What are practical recommendations for teachers and designers?</p> | <p>Q4</p> <p>Are these abstractions also useful when designing for sustainability?</p> <p>a) Is the CLOVER framework applicable in different settings?</p> <p>b) What design factors are critical for achieving sustainability in the wild?</p> |
| Study | 4Decades (lab) | 4Decades in the wild | | UniPad Comfy Birds |
| Data | lab video, field notes system logs teacher interviews student questionnaires debriefing transcripts | in-situ video, field notes system logs, screenshots teacher interviews student questionnaires debriefing transcripts | findings from 4Decades and literature | notes from design meetings with teachers teacher interviews field notes system logs |
| Analysis methods | lab-based prototyping qualitative analysis | detailed video analysis qualitative analysis of system screenshots | reflection on design development of a conceptual framework | lab based prototyping preliminary feasibility evaluations in the wild qualitative analysis |

Figure 4.1.: This chapter in the context of other chapters.



Figure 4.2.: The 4Decades simulation for two teams and a teacher, using shared displays and tablets.

4.1.1. Research question and aims

The research question that guided this particular chapter was:

Q1: How can whole classroom activities be designed that use shared technology to facilitate large groups' problem-solving and collaboration in a face-to-face setting?

- a) What types of engagement goals are desirable and feasible?
- b) What mechanisms can effectively support these goals?

The following research aims were addressed.

1. to explore feasible design options regarding (1) pedagogical integration, (2) social mechanisms of collaboration and facilitation and (3) devices and displays;
2. to implement a prototype application that can be used in a real setting.

4.1.2. Chapter Structure

This chapter begins by explaining the context of the study, the setting and overall research approach (Section 4.2). It goes on to report on an initial ethnographical investigation (Section 4.3) that informed the objectives for the design. These objectives are then summarised and used to devise some preliminary choices of design methods and starting points (Section 4.4). Three sketches are then presented, each in its own section (Sections 4.6 - 4.8) describing the design, evaluation and findings. The chapter concludes with an overall discussion of the findings (Section 4.9).

4.2. Study context and approach

This section explains the context of the study, i.e., a particular series of executive training courses, and the overall approach taken. The latter involves a description of the concrete learning need, co-operation with the course committee, the teacher's role in the design and the time frame of the project.

4.2.1. Executive training courses

The 4Decades study was conducted in the context of a series of executive training courses organised and hosted by the Cambridge Programme for Sustainability Leadership (CPSL)¹. This institution has a long tradition of aiming to deliver high-quality, executive training. According to their website, their programmes have *“two distinctive features: the quality of the discussion and debate; and the extent to which we challenge and stretch participants to think more broadly and longer term about their own leadership role and the resilience and growth of their organisation.”*² As this statement testifies, the programmes focus on higher-level learning and discussion, rather than mere information outlets. For this purpose the programmes have included not only lectures but also active learning components, such as workshops, case studies and simulations.

4.2.2. Addressing a concrete learning need

A concrete, pedagogical need was expressed in regard to one of CPSL's programmes - particularly, the Climate Leadership Programme (CLP) that aimed to increase senior executives' knowledge, skills and confidence to improve their organisation's policies with respect to global sustainability. In pursuit of this goal, the courses provided lectures and workshops on climate science, economics and policy making.

For several years the course committee had been struggling to provide an adequate activity for participants to apply theory from the course lectures in practice. Off-the-shelf solutions had previously fallen short of the goal to efficiently engage a whole cohort of 20-30 participants and encouraging high-quality discussion of very specific, systemic aspects of climate and sustainability. This suggested a novel design challenge with potential significance for HCI, game design and collaborative learning.

4.2.3. Research in practice: co-operation with the course committee

The co-operation with CPSL was agreed as a research project, rather than a paid design job, based on the following main premises.

¹Now Cambridge Institute for Sustainability Leadership www.cisl.cam.ac.uk

²www.cisl.cam.ac.uk/About-Us/Our-Approach.aspx retrieved on 5th August, 2014.

- CPSL agreed to let us evaluate a prototype in one of their courses, provided that the prototype was technically robust and the activity met their pedagogical requirements.
- In return for the activity's educational value, I was given permission to collect data during the courses within reasonable constraints.
- Furthermore, it was suggested that our activity - if deemed superior - might become a regular part of the course series. This indicated potential for future evaluations, which in turn would benefit our research by allowing design iterations and more rigorous validation of our approach.

The academic work - including the literature review, design, implementation and evaluation - was conducted by myself, using my advisors' help as needed. CPSL generously assisted by sharing their pedagogical experience, giving feedback on our sketches and prototypes. Moreover, they supported the evaluations by providing the rooms, furniture and some of the technical equipment, such as projectors.

A requirement was that the activity - besides its research purpose - also had to satisfy participants' expectations of a highly efficient learning experience. This implied that the activity should be engaging, enjoyable and useful for all participants. Given the workshops' reputation and high cost to participants, there was little tolerance for poor delivery or technical failures that would likely cause distractions, confusion or waste of participants' time. My responsibility was to ensure that the application was robust in all respects (technically, pedagogically and socially) before taking it out into the field. This required extensive testing with volunteers in the lab.

4.2.4. Whole-activity testing in the lab

Each lab evaluation was conducted with a focus on the activity as a whole. In contrast to controlled lab studies that aim to "*isolate the effects of particular factors on performance, such as size or orientation of interface*" (Rogers et al. 2013) the focus here was on testing the application as a whole, to see whether it breaks; if so, when, where, how and why.

For this purpose, lab participants were thrown into a joint problem-solving situation, very similar to the real target setting.

4.2.4.1. Replicating the wild in the lab

An effort was made to replicate as many aspects of the ultimate use scenario as feasible. This included using the same technology, the same teacher, a similar population (highly educated and polite adults) and a similar overall procedure (introduction, simulation, debriefing). To minimise the risk of unexpected issues in the wild, the following steps were taken:

1. Rather than isolating specific factors, the activity was tested as a whole to identify any unexpected bottlenecks, critical incidents, breakdowns and breakthroughs.
2. Participants were recruited from a population with typically high analytical and communication skills who were able to raise constructive questions and comments. Their help was appreciated in regard to analysing emergent issues and predicting potential risks in the real setting.

4.2.4.2. Participant feedback and observations

Lab participants were encouraged to bring any difficulties, concerns or suggestions to the researchers' attention. This implied creating a friendly and informal lab situation in which participants were unafraid to ask. Moreover, an effort was made to elicit feedback regarding all Triple-i layers. In addition to participants' explicit reports in the debriefing, the evaluation relied on the researcher's observations of groups interacting.

4.2.5. The teacher's role in the design

Stephen Peake played a central role in this cooperative project. Being a member of CPSL and regular speaker at CLP courses, he brought an instrumental body of experience to the project. This included his expert knowledge on global climate issues, familiarity with CPSL's didactic tradition and the specific course pedagogy, a deep understanding of the target population (particularly attendees' professional backgrounds, learning needs, abilities and dispositions), as well as relevant skills of lecturing and facilitating group work. Simultaneously, being involved in this research project as a thesis co-advisor and co-author on all the related publications, he was also comfortable with aspects of interaction design and constructivist pedagogy. For these reasons it was decided early on that Stephen would be responsible for introducing, facilitating and debriefing the activity.

4.2.6. Time Frame

The project began in October 2010 when Stephen and I met with leading CPSL members in Cambridge to elicit requirements and outline potentially viable approaches. The next upcoming course event had already been scheduled for end of January 2011. For logistical reasons CPSL insisted on knowing one month in advance whether our team would be able to deliver a feasible solution on time. This gave us less than three months to design and develop a working prototype that could meet all the requirements.

4.3. Study setting: The CLP course series

To elicit requirements, a number of steps were taken to gain a holistic understanding of the training course in its pedagogical, physical and organisational context. This included:

1. the overarching pedagogical context, i.e., the course curriculum and the organisation's general teaching philosophy;
2. participants' backgrounds and specific learning needs;
3. what efforts had been conducted previously to fulfil these needs - what approaches had worked well and what didn't;
4. constraints of the physical setting and how new design might potentially improve the situation.

Some of this information could be gained directly by interviewing Stephen who had been part of the course committee and thus well informed, particularly with regard to points (1) and (2) - to some extent also point (3). To further our understanding of point (3), Stephen and I met with other members of the course committee. Ultimately, I was invited to visit part of a training course to get some first hand experience with the type of physical and cultural environment in which the courses were situated. This was essential for point (4). The outcomes of these investigations are summarised in the four subsections below.

4.3.1. Motivation for active learning

According to the members of the course committee we talked to, the rationale for having a role-play activity at this point included a number of key pedagogical intentions.

1. **Allow attendees to apply learned theory in practice**, i.e., to engage actively with the specific course content as part of an activity, to support higher-level learning, deep understanding and retention.
2. **Refresh the tired group, mentally and physically**, by switching to a different form of engagement - one that encourages playfulness, humour and body language. This was further intended to “break the ice” between participants, i.e. to reduce social discomfort and encourage dialogue. Considering that some of the attendees represented competing organisations or otherwise conflicting interests, this was considered one of the main goals of the activity.
3. **Create a collective experience** that subsequent discussions and workshops could reflect on.
4. **Help faculty gain insight into participants’ learning**, by letting them verbalise their beliefs and assumptions during the activity.

Considering the substantial course fees, a high quality was expected of the course as well as the role-playing activity.

The four-day course schedule was organised in a lectures-first manner, reflecting a pedagogical strategy of first establishing basic theoretical knowledge and then allowing attendees to apply this knowledge in hands-on activities. Since attendees are usually very tired after two days of mostly lectures, a slot in the afternoon of day two was reserved for a role-playing game or simulation (see Figure 4.3).

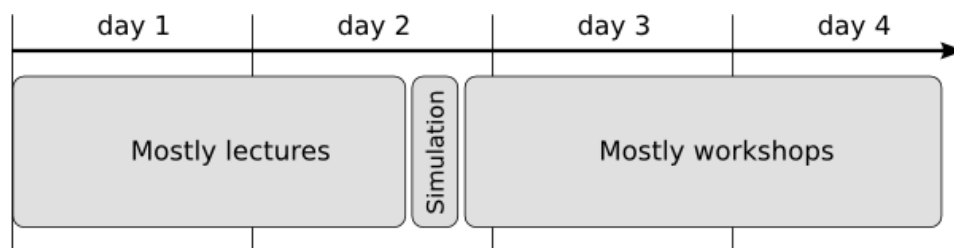


Figure 4.3.: A typical CLP course featured four days of lectures and workshops, with a two-hour slot for a simulation in the afternoon of day 2.

4.3.2. Participants' educational backgrounds and needs

In regard to the entire user population, we had a fairly clear picture of what to expect: A closed group of about 30 university-educated professionals with high communication skills, all fluent in English and familiar with standard office technology. We knew that their backgrounds ranged from engineering to business and public relations and that hence some of them would have advanced mathematical skills, leadership qualities and other skills, respectively. Most importantly, we knew the exact content of all the course lectures and the course schedule. This allowed us to know what background knowledge to take for granted when designing the activity. We knew that they would be reasonably tired after those lectures and their heads would be full of new information. Finally, we knew that the group would come as a whole and leave as a whole.

During the course, attendees had very little time to spare, as they were expected to attend the full programme and use the breaks for socialising and networking.

4.3.3. Previous efforts towards applying theory in practice

The “active learning” slot of the course series had previously been filled with sustainability-themed board games and simulations. Numerous off-the-shelf products had been tried and considered by CPSL over the years. For instance, the FishBanks simulation (Meadows et al. 1989) was used to illustrate the “tragedy of the commons” principles by allowing participants (in the roles of fishers) negotiate about how to use a shared, renewable resource (in this case, where to fish, how much, etc). More recent courses had used the Perspectivity board game (www.perspectivity.org), due to its topical relation to climate. However, none of the available options were deemed completely satisfactory for the purpose.

- Board games were appreciated for their face-to-face properties but considered problematic, due to their lack of numerical accuracy and their inability to accommodate 30 simultaneous players. Since board games are limited to a small number of players, previous uses of board games ended up fragmenting the group into multiple, isolated boards (tables) that did not interact with each other.
- Dennis Meadows' classic FishBanks simulation takes at least three hours to explain, conduct and debrief (Meadows 1999) which was considered too long in the context of the training course.

- Sophisticated climate simulations, such as the C-ROADS model developed at MIT (Sterman et al. 2013), appealed to the course committee due to their accuracy and immediate feedback. However, the question of how to give 30 people simultaneous access to the simulation remained, since C-ROADS is based on a single-user (web browser) interface that only one person can interact with at a time. The prospect of 30 delegates sharing a mouse and a keyboard did not appeal as a promising road to go down - neither did the possibility of several workstations at which individuals or small groups would work in isolation.

Concluding from the literature review and CPSL's prior experience, it was safe to assume that board games, desktop applications and paper-based simulations could be ruled out, due to their inherent limitations, including time demand and lack of shareability.

Besides the above issues of multi-user accessibility, according to CPSL's experience, many off-the-shelf simulations "send the wrong message", i.e., misled participants into false conclusions, due to distorted or poorly simplified representations of the problem area. For instance, the Perspectivity game was criticised for promoting a highly contested notion that emissions have no cumulative effect and thus policy makers need not worry about global warming as long as emission levels stay below a constant threshold. The committee emphasised how problematic such misrepresentations were for the overall learning goals, as they contradicted the previous lectures in a less than helpful way. In other words, scientific accuracy was highlighted as an important design goal for the simulation.

4.3.4. Ethnographical investigation

I was given permission to attend part of a CLP course and participate in a group activity (the Perspectivity game). By immersing myself in the concrete training situation, I aimed to achieve two things:

1. to gain some relevant insight into the physical, organisational and pedagogical constraints;
2. to identify opportunities for a technological solution to improve the situation by making the experience more exciting, enjoyable and importantly, putting into practice the theory that was taught in the course lectures.

4.3.4.1. Starting points for interaction design

The activity I participated in was a traditional board game, of which three instances were played in parallel on three large tables. Each table had 12 participants and one dedicated facilitator. Although the game was thematically related to climate change, its promoted values and abstractions somewhat contradicted what was taught in the lessons. Furthermore, the course organisers were unhappy with the way that the game divided the group into isolated sub-groups, rather than creating a sense of whole-group togetherness. The game itself could not provide this experience, despite having human facilitators walking around in the room. It was hypothesised that technology might fulfil a vital role in this regard.

4.3.4.2. Starting points for interaction design

My observations raised a number of challenges concerning how to design an application that all could take part in together in a classroom setting. For instance, if the activity were to rely on interactive tabletops, then at least 4 tabletops would be needed, which was not feasible or practical to set up for a 2-hour session. Therefore, alternative options had to be considered, such as laptops, projectors, tablets or mobile phones. Moreover, those devices probably had to be networked in some way to encourage awareness and communication across the whole group. Finally, the educational content of the application would need to fit in with what the group had been taught, which suggested that the design process should involve a domain expert with a deep understanding of the course.

4.3.4.3. Eliciting the language of teachers: The notion of “surfacing”

Engaging with the course committee sharpened my understanding of their pedagogical goals. Although what they said in conversation was - with regard to content - consistent with the official, written course descriptions, the internal conversations revealed some additional insights into their practical approach to teaching. One thing that particularly struck me was the metaphor of “surfacing”, in the sense of *bringing to the surface* participants’ tacit knowledge, beliefs, assumptions and biases (Figure 4.4).

With CPSL’s pedagogic tradition, surfacing is considered instrumental to critical reflection and empathy with conflicting opinions. It is best described as a guiding principle and metric for successful teaching. It is characteristically applied to collective group processes,

rather than individuals. Furthermore, it is typically considered to require a conscious effort from the teacher, such as stimulating questions or guided debate.

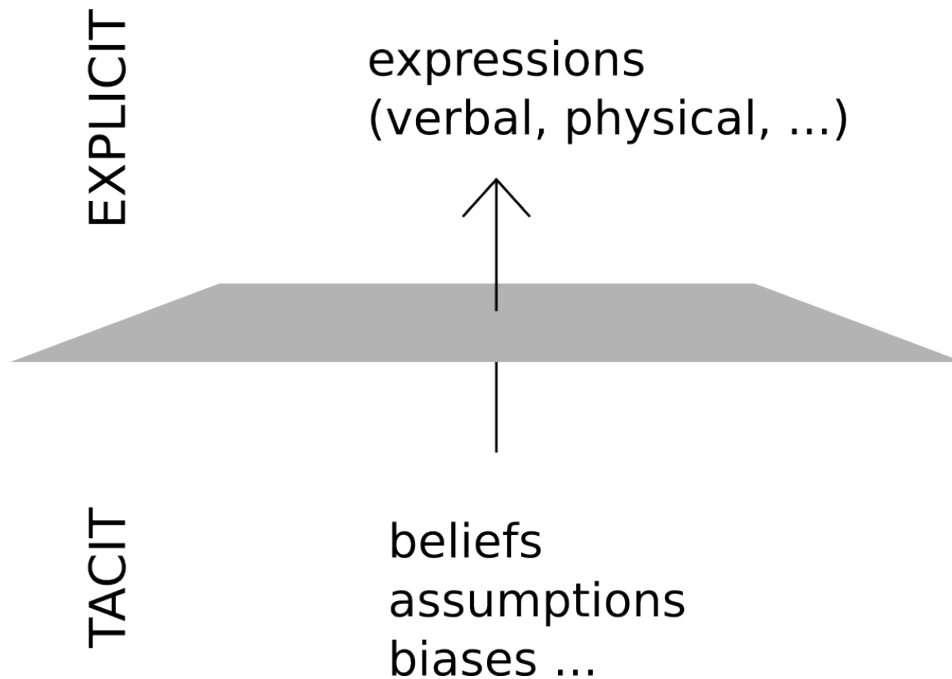


Figure 4.4.: One goal of pedagogy is to “surface” tacit understandings.

The expression “surfacing tacit knowledge” has been commonly used in practice-oriented fields, such as management (e.g. Ambrosini 2001, Bennet 2008) and requirements engineering (e.g. Gacitua 2009). Variants of the same principle can be found in many learning theories. For instance, Stahl’s (2006) model of collaborative knowledge construction mentions “tacit pre-understandings” that are “made problematic” (Figure 4.5). Although arguably these scholarly terms are useful in the context of the author’s fine-grained analysis, they may be less usable at the level of committee meetings. That is where concise metaphors such as “surfacing” are sometimes preferred.

It is worth noting that this particular use of the “surface” metaphor must not be confused with other uses of the “surface” metaphor. For instance, Moss et al. (2007) distinguished between “deep understanding” and “surface understanding” - the latter implying that someone (in this case a teacher) may not be fully aware of the potential and ramifications of something (in this case the use of whiteboard technology). Moreover, the word “surface” is frequently used as a label for interactive touch screen technology, e.g., the Microsoft Surface or a workshop series titled www.surfacelearning.org.

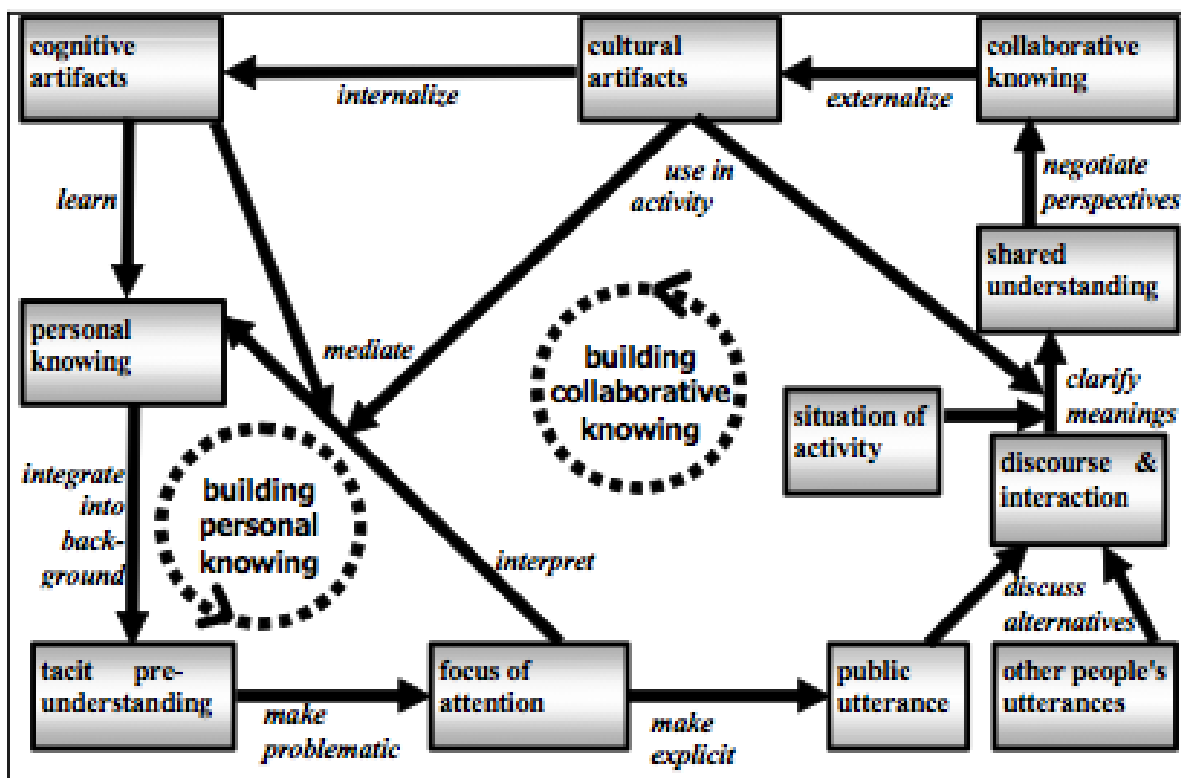


Figure 4.5.: Stahl's (2006, p.313) model of collaborative knowledge building

4.4. Design method

This section first presents the design objectives that emerged from the meetings and field observations. I refer to them as “Gameplay Objectives” because they focus on social activity during the simulation or game. Afterwards, this section discusses a number of choices regarding topics, evaluation strategies, activity structures and appropriate technologies.

An iterative design approach was chosen, using the Triple-i framework (instruction, interaction, interface - see Chapter 3) as a lean guiding tool for design and evaluation. Following the priorities suggested by Triple-i, each sketch or prototype was designed by first narrowing down on the instructional goals - including the exact “topic” (content / lesson) of the simulation, before considering the type of group interaction and finally HCI factors. The same sequence was applied to evaluation.

4.4.1. Gameplay Objectives

Following the ethnographical investigation, a number of design objectives were formulated to support whole-class gameplay. These needed to be achieved in order for the activity to be considered a success. To allow rigorous evaluation, they were phrased as specific objectives (rather than mere ideals) that could be verified or falsified through observation.

1. Allow participants to apply relevant course knowledge in practice through decision-making and debate;
2. Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use;
3. Allow a single teacher to facilitate and debrief the activity, using an ecology of displays;
4. Encourage participants to reflect on assumptions through decision making and discussion;
5. Allow participants to refresh from exhausting lectures by moving and talking;
6. Encourage negotiation by letting participants follow different goals but requiring the consent of teammates;

7. Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play;
8. Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data;
9. Allow all participants to contribute equitably to the activity and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics.

Embedded in these 9 objectives are aspects that concern all three layers of the Triple-i hierarchy. Six of these objectives concern aspects that span several layers, with a one-way dependency implied. These cross-relations are represented by the arrows in Figure 4.6. It is worth noting that none of the arrows points to layer 3, indicating that the technology was always seen as a means to an end, rather than a goal. Furthermore, the figure betrays a strong focus on i2 where the main problems of feasibility are assumed to be located. The arrows pointing to i1 testify that the design effort is ultimately directed towards instructional needs.

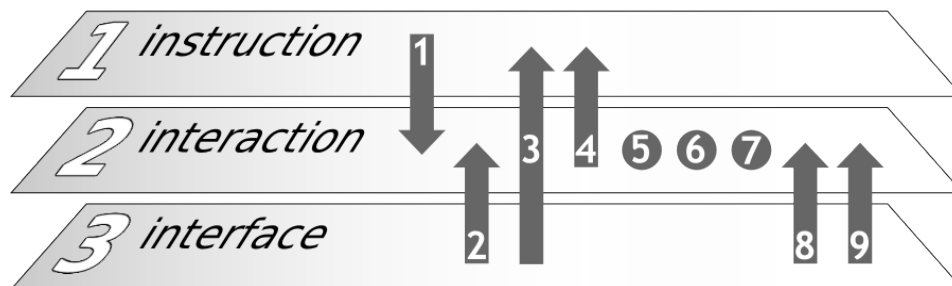


Figure 4.6.: Mapping the 9 Gameplay Objectives onto the Triple-i layers

4.4.2. Choice of topics

Topics were chosen according to the following three pedagogic principles, which required that the topic

1. was of high relevance to the course programme, rather than merely tangentially related;
2. involved aspects of human decision making, climate and sustainability;
3. could not be taught adequately through lectures alone.

Besides these pedagogical criteria, there was also the *practical* consideration that the topic should lie within the subject-matter expertise available in our team. This would allow

the team to remain small and versatile, rather than relying on external subject-matter experts.

4.4.3. Choice of evaluation strategies

The first sketches were designed based on initial assumptions and evaluated by discussing them in our research group. Estimates about the strengths and weaknesses of a design relied strongly on our previous experience in HCI and collaborative learning as well as Stephen's concrete experience as a facilitator and lecturer in CPSL courses. Later iterations required interactive prototypes that were tested with large groups in the lab. After three cycles of design and testing, the activity was considered mature for testing in real training courses.

4.4.4. Devising an activity structure

The setting of the conference centre allowed to use multiple rooms, hallways, etc. if necessary. Constraints were that it had to take place indoors after a specific set of lectures. The course schedule required that the activity take less than two hours.

As a first educated guess, the design aimed to combine the benefits of participatory simulations with a team game, involving strategy, collaboration and competition. This form of play was believed to produce many affordances for exciting social interaction. Moreover this form of play was considered a broadly familiar, popular and accepted technique within our culturally mixed target group.

4.4.4.1. Giving participants a real problem to solve

Perhaps the biggest motivation for using teams in this context was due to the particular topic of climate change. Due to climate change being a 'wicked problem' (Rittel & Webber, 1973), it would be presumptuous of any designer to predefine an absolute winning condition. Having participants compete against each other is a way for designers out of this dilemma - since competitive games only require a *relative* winning condition. In other words, we don't have to quantify the meaning of "good enough" - only that of "better". In the context of climate change education, this is still far from trivial but it is manageable.

4.4.4.2. Influences from previous work

Previous research in collaborative interaction design suggested further potential to engage participants in fruitful social interactions by carefully constraining information, control and awareness (Yuill & Rogers, 2012). For instance, using multiple linked representations (Lyons et al., 2006), public and personal territories on large touch screens (Scott et al., 2004) and supporting opportunistic, over-the-shoulder learning (Twidale, 2005).

Participatory simulations

Since the CLP course focused on leadership and decision making, it seemed pedagogically conclusive for participants to experience the simulation from the perspective of decision makers - as opposed to, say, outside observers. Given the large group size and the goal to engage participants face to face, the participatory simulation format seemed to be the most promising starting point for the design.

Inspiration from embedded phenomena

Further inspiration came from Moher's work (2006), showing how in a room-wide simulation, participants can opportunistically take on different roles and share a variety of task, ensuring that everyone has something to do. For instance, in the RoomQuake paper, Moher writes:

“While the number of electronic affordances was less than the number of students, the additional roles (twine puller, magnitude computer, supervisor, etc.) were sufficient to occupy a class of 19 students.” (Moher, 2005)

Conceivably, this principle might also work with a wider range of tasks, not necessarily limited to Moher's examples - notably scientific observation, measuring, hypothesising, etc. - but also potentially other activities, such as information foraging (Pirolli and Card, 1999), collaborative search and sense-making (Morris et al., 2010) and negotiation. To keep a large number of participants occupied, give them different roles was initially considered a promising option for this design.

Inspiration from role-play and games

With debate-oriented role-play activities it is common that participants are divided into groups that represent populations with different interests, to help participants understand conflict and develop empathy through discussion. Morgan (2003) argues that such simulations can highlight (a) why conflict is so prevalent and (b) how difficult global cooperation can be to achieve.

This notion of developing empathy by negotiating conflict was considered highly relevant to the CLP course. Guided by this idea, early sketches capitalised on several participant teams confronting each other in a situation of conflict. Later versions, however, favoured a kind of competitive team play without direct conflict, to avoid mutual obstruction and domination. Competition - or rather, mutual comparison of results - was used as a social mechanism to increase motivation and stimulate ideas for different strategies.

4.4.5. Choice of technology

As different types of devices imply different affordances and constraints, it was hypothesised that by combining adequate types of devices into well-constrained ecologies, the design could be shaped into supporting the desired types of individual and social interactions. The eventual decision to use tablets and shared displays was made after initial explorations had found that those would best support the type of group interaction we ended up aiming to support - i.e., open discussion about semi-private and public information, with equal roles for all participants. Arriving at this result required several design iterations. Three lab-based design iterations are outlined in the following sections.

4.5. Introducing the three sketches

The following three sections describe the design work that was conducted in the lab prior to in-situ deployment. Each section represents a significant step in the iterative design process and follows the same structure: Design, Evaluation, Discussion and Summary. To describe the design changes and guide the analysis and discussion, the Triple-i framework is applied. Particularly, each of the “Design” subsections contain dedicated paragraphs on aspects of instruction (i1), interaction (i2) and interface (i3). The “Discussion” subsections refer back to these categories accordingly.

4.6. First sketch: “Table of leaders”

Below I report on the first idea that was developed into a somewhat elaborate sketch. The sketch was titled “Table of leaders”, because it envisioned a table in the centre of the room where a small group of participants represented their teams as leaders (see Figure 4.7). Although this design was not followed through, it is worth reporting, due to its rationale and the questions it raised.

4.6.1. Design

Adhering to the Triple-i hierarchy, the design primarily focused on the key instructional intention before considering subsidiary interaction factors and interfaces.

4.6.1.1. Instruction design (i1)

The sketch focused on aspects of leadership and communication, particular the following real-world problem:

International negotiations, including the 2009 Copenhagen Summit, have suffered from the fact that delegations were unable to communicate easily with their leaders in real time.

In line with the course topic, the simulation was intended to help participants gain an experiential understanding of the following questions: What is it like for delegates to be in such a situation? What are the implications for decision making?

4.6.1.2. Interaction design (i2)

Aiming to replicate this situation in the physical space of the conference centre, a distributed activity structure was devised.

The idea was to divide the group into reasonably small teams - e.g., 4-6 participants each - that represent countries on a planet, similar to the Perspectivity game. Each team would initially elect a leader (or ‘head of delegation’ or ‘negotiator’) who would negotiate at a central table face-to-face with the other teams’ leaders. The remaining team members had the roles of “analysts”.

As illustrated in Figure 4.7, leaders could exchange messages with their analysts (e.g., using handheld devices) that were hidden from the other teams.

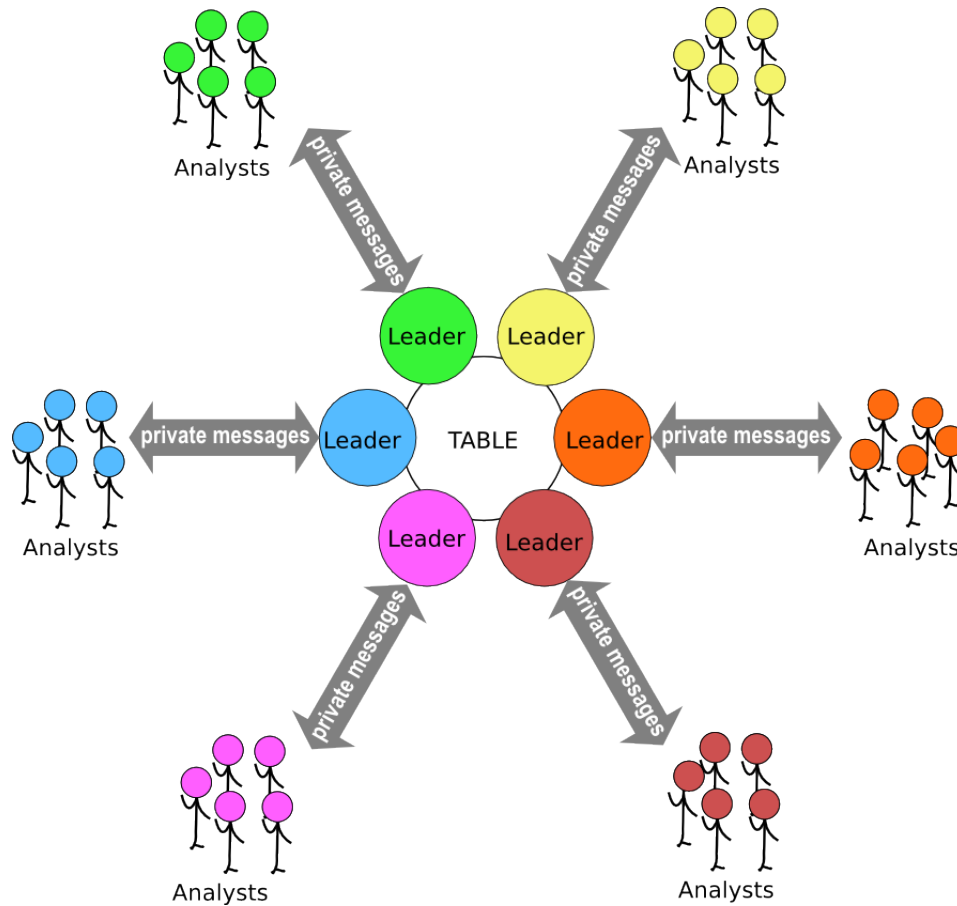


Figure 4.7.: Schematic illustration of “Table of leaders”, with multiple teams (colours). Team leaders could talk face-to-face at their table whilst also exchanging private messages with their analysts.

The activity structure can be summarised as follows:

- Teams are given competing goals (winning conditions), resulting in a conflict of interests between the leaders at the table;
- Analysts and leaders receive complementary information about the simulation's state, creating an incentive to communicate;
- Communication between leaders and analysts is two-way. They can exchange private messages but not talk face to face;
- The private message channel is designed as a bottleneck, i.e., a deliberate constraint, requiring participants to figure out efficient strategies for exchanging relevant information quickly;
- The table could provide specific information that all the leaders would get, but not the analysts, e.g., using a shared horizontal screen;
- Decisions must be made in multiple periods, e.g., game rounds or fixed time intervals;
- Either the leaders or the analysts should have some kind of input device for entering their team's decision at the end of each decision-making period.

This overall configuration was expected to create much potential for serious on-topic negotiation as well as socially playful game mechanisms, such as bluffing, double bluffing or back stabbing.

4.6.1.3. Interface design (i3)

The development of this sketch was discontinued before it reached the point of finalising decisions about the type of devices and displays and the design of interfaces. An example schematic (Figure 4.8) was created merely to illustrate the activity. The depicted devices - laptops, handheld devices and an interactive tabletop - should be seen as placeholders for a range of possible devices and interfaces.

4.6.2. Evaluation

To make initial ideas more concrete and specific, I created a brief design proposal with some diagrams, as if explaining the idea to someone external to the team (e.g. the course organisers). This proposal remained in a crude draft stage and was never presented

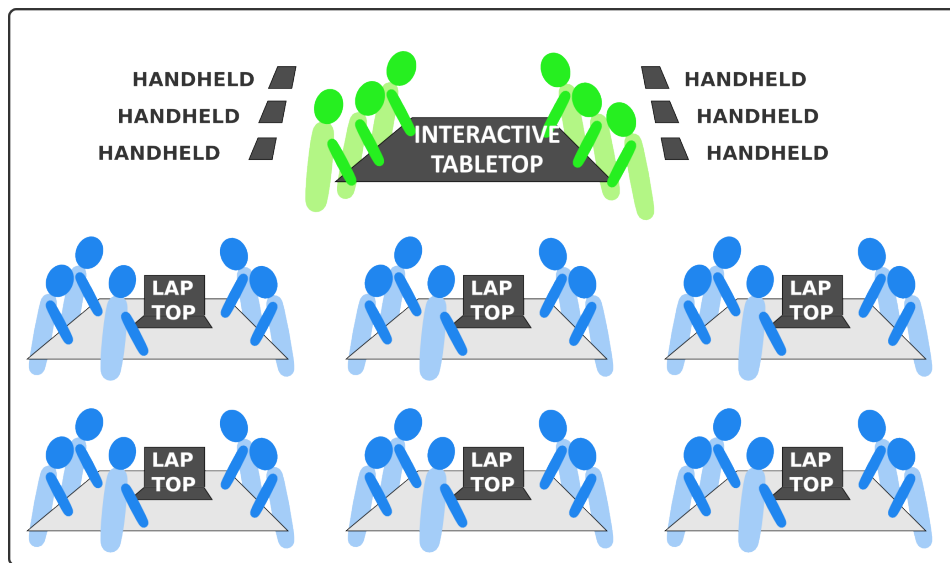


Figure 4.8.: Leaders (blue) and analysts (green)

or even elaborated in detail, since discussing the first draft with HCI experts already produced enough insight for revising the design. Particularly, the following questions and potential shortcomings were identified.

Instruction (i1)

1. How could the activity be debriefed as a whole if participants at different tables experience essentially separate activities?
2. Singling out a minority of “leaders” defies the course’s goal to create a situation for attendees to overcome organisational, cultural and political boundaries. In this sense the design would risk provoking undue social imbalance and discomfort within the group, especially if one of the roles was perceived as more interesting, fun or useful for learning.
3. Assigning multiple roles (i.e., leaders and analysts - as opposed to giving everyone the same role) appears to have further implications:
 - More time would be required for introduction and explaining the rules.
 - Joint debriefing may be compromised due to the fact that leaders may not have experienced what the activity was like from the analysts perspective, and vice versa.
 - Conceivably, playing one role or the other might potentially result in different learning outcomes. This might diversify the group and potentially impact subsequent course activities.

Interaction (i2)

1. If each subgroup does not know what the other subgroups are doing, how could the subgroups learn from each other?
2. How could a group of analysts equitably divide their labour (e.g. information foraging, messaging, etc)? Would one analyst be likely to end up doing all the work while others are excluded or freeloading?
3. Splitting the group into many small teams may be detrimental to the goal of encouraging dialogue among *all* participants.

Interface (i3)

1. How could the team leaders protect their private (on-screen) information from other teams, without it becoming physically awkward?
2. What would be appropriate interfaces for the analysts to share equitably?

4.6.3. Discussion

The “Table of leaders” sketch represented a first educated attempt at marrying a relevant course topic with a promising activity structure. Creating and analysing this sketch raised some important questions that informed further design revisions. These questions are discussed below, using the Triple-i hierarchy to structure the discussion into aspects of instruction, interaction and interface.

4.6.3.1. Instruction evaluation (i1)

The topic chosen for this sketch was about communication bottlenecks during climate negotiations. Although this topic fulfilled the requirements of being relevant to the course and potentially benefitting from a role-play perspective, it was found to entail a host of i2-related issues that conflicted with some of the overall design goals.

Since these conflicts seemed severe and difficult to resolve, it was considered more pragmatic to find a different topic.

Based on the above considerations, a more suitable topic would allow for participants to all have the same roles or be able to swap roles easily.

4.6.3.2. Interaction evaluation (i2)

This design emphasised *hotspots* of group activity that were distributed in the room, based on computational affordances and team structures. By analysing the sketch using the Mechanisms for Collaboration framework (Yuill & Rogers 2012) a number of collaboration-related problems were highlighted (a) within the hotspots, (b) between the hotspots. While the problems in (a) seemed essentially resolvable, (b) contained some aspects there were considered real show stoppers.

Within-hotspots collaboration

(1) Mutual awareness of actions and intentions. Since the team size was chosen rather small, awareness within teams was considered the least problematic of all the issues raised in this context.

(2) Control over the interface and (3) availability of background information. Although the sketch did not propose any concrete solutions in these regards, a range of options would have been reasonable to explore, including laptops, tablets and handhelds.

Between-hotspots collaboration

(1) Mutual awareness of actions and intentions. Constraining this factor was a key intention of the sketch, in order to increase difficulty and encourage teams to find clever workarounds by collaborating. Although the sketch proposed some ideas, such as private text messages, these were seen as problematic. Most importantly, the need to type private messages would distract participants from face-to-face talk and tie them to individual screens. This conflicted with number 5 of the 9 Gameplay Objectives, related to encouraging talk and moving around.

(2) Control over the interface. The idea that leaders and analysts would use different and separate interfaces to control the simulation was considered very problematic with respect to several Gameplay Objectives (at least 2 and 9).

(3) Availability of background information. The sketch proposed that leaders and analysts are given complementary background information about different parts of the simulation. This violated objective number 8, i.e., to encourage mutual awareness and extended dialogues among peers. Furthermore, the separation of knowledge between leaders and analysts suggested likely negative implications for facilitation and debriefing (objective number 3).

In summary, the above problems were considered severe and inherent to the main idea of the sketch, thus requiring a fundamental redesign.

Hotspots in related work

It was considered that hotspots are per se a useful thing, as they can help structure an activity, but that the current sketch used them in a less than optimal way (from an interaction perspective) by imposing artificial boundaries between the hotspots. In search of better solutions, a number of related studies were analysed with regard to hotspots and their interrelation. Three recurring themes were discovered: (a) *spatial mapping* of locations in the classroom to locations in the simulation, (b) *task-specific affordances* and (c) *opportunistic mobility* of participants between hotspots.

Spatial mapping was found in several of the embedded phenomena, including RoomQuake (Moher, 2005) and AquaRoom (Novellis & Moher, 2011) and some augmented-reality simulations, especially outdoors activities (Klopfer et al., 2002; Benford et al., 2004).

Task-specific affordances: by this I mean that the a hotspot was mainly constituted by the type of task it allowed participants to perform - following the same principle as a vending machine in a large hall or a popular slide on a playground. For instance, workstations represented virtual food patches for foraging animals (Gnoli et al., 2014); Flatscreen monitors attached to the walls allowed students to peek into the hidden space within the classroom walls (Moher et al., 2008). Such a task-specific distribution of affordances (unlike the role-based distribution proposed in the given sketch) allows participants to use one hotspot or another based on what they currently need to do. These needs may be more or or less emergent and dynamic, depending on the structure of the activity. A good design should avoid excessive crowding and the need for participants to queue.

Opportunistic mobility: Moher (e.g. 2005) emphasises the benefits of participants moving between hotspots, opportunistically switching roles and thus passing on experience to their peers through collaboration.

In the “Table of leaders” sketch the idea of physical and computational *hotspots* was quite pronounced. However, participants were tied to *their* particular hotspot based on their role and team. Arguably, this lack of mobility and exchange compromised the potential for community-of-practice learning. This was considered a severe shortcoming of our design that needed to be addressed in the subsequent iteration.

Based on the above considerations, the following suggestions were made for the next design iteration:

- There should be only few roles, to save explanation time and facilitate orchestration;
- If there is more than one role, participants should be able to switch between roles;
- Swapping roles should be encouraged, either through some explicit rule or indirectly via mechanisms. In either case, participants should be granted sufficient time in each role as relevant for their learning;
- Teams should be able to collaborate freely without any imposed restrictions;
- All team members should be allowed to control the simulation through the same kind of interface;
- All team members (and potentially non-members, too) should have access to the same background information;
- Communication *between* teams need not be regulated;
- Depending on the topic, it may be beneficial to distribute hotspots according to some meaningful spatial mapping;
- Hotspots should derive their primary function and meaning from the tasks they enable, rather than the groups or individuals they belong to;
- The design should allow participants to use the hotspots as needed, while avoiding congestion.

4.6.3.3. Interface evaluation (i3)

The Triple-i hierarchy suggests deferring concrete interface decisions (i3) until the interaction goals (i2) are precisely specified. Therefore, since the i2-suggestions listed above were rather broad (and partly tentative) no concrete suggestions were made regarding interface design at this stage.

4.6.4. Summary

Designing and evaluating the sketch according to the Triple-i hierarchy suggested three notable benefits:

1. The sketch was based on a solid instructional rationale, in the sense that the topic was highly relevant to the course.
2. Shortcomings in the interaction design were detected early and raised important questions.
3. No time was wasted on designing interfaces, let alone programming.

In summary, the “Table of leaders” idea was considered a good start, despite considerable problems in the interaction layer that led to its discontinuation. The next version aimed to address the issues and questions raised.

4.7. Second sketch: “Parallel futures”

Based on the previous findings, the design goals were refined as follows:

- The design should not restrict verbal communication;
- To encourage broad communication, participants should be allowed to move and mingle. This does not exclude the idea of hotspots, but participants should not be tied to those in a static way;
- Small team sizes are problematic to the extent that they restrict mobility and broad communication;
- It is preferable for the design to use fewer roles, in order to reduce explanation time and let participants have comparable experiences that can be jointly debriefed;
- In fact, all participants should have *one and the same role* to support easy conversation and equitable participation among peers;
- In absence of multiple roles, alternative mechanisms are needed to facilitate participants’ division of labour.

The above goals suggested substantial changes to the interaction layer (i2). Furthermore, there was an obvious, irreconcilable contradiction between these new goals and a topic that relied on a communication bottleneck. Therefore a different topic had to be found.

4.7.1. Design

As with the other sketches, the Triple-i hierarchy was used to prioritise the design considerations.

4.7.1.1. Instruction design (i1)

The new sketch, entitled “Parallel futures” was based around a newly formulated question:

What will be the cost of anthropogenic climate change in the next 40 years, and how should policy makers balance investments between greenhouse gas mitigation, adaptation to climate change and residual repairs?

This question was chosen due to its key relevance to both the course curriculum as well as participants’ professional situations as policy makers in international organisations.

4.7.1.2. Interaction design (i2)

The interaction layer was redesigned from scratch to accommodate the revised design goals. To encourage broader communication, the new sketch envisioned two large teams, rather than many small ones.

Two equal planets

In contrast to the previous sketch, where the teams represented countries on a shared planet, this version took a different approach: Each team had its own, earth-like planet and made independent spending decisions about CO₂ mitigation, adaptation and residual repairs. The winner was defined as the team with the “healthiest” planet at the end of several decision rounds.

One table for each team

This version emphasised that teams could learn from each other, e.g., by comparing outcomes, imitating or improving strategies or stealing information. For this purpose, the design featured two long tables, one for each team, assuming that this would allow participants to discuss easily within their team and also with the other team. Each team was given a reasonable number of computational affordances for entering their decisions into the shared simulation. For instance, four devices per table as illustrated in Figure 4.9.

An element of time travel

To give the simulation a unique twist, it was suggested that the devices could represent years on a timeline, e.g., 2010, 2020, 2030 and 2040 as suggested in Figure 4.9. Allowing present and future inhabitants of the planet to make climate decisions face to face. Using a distributed model of climate change, participants attending, e.g., the year 2040, could immediately see the effects of decisions that their teammates, e.g., people in 2010, are making, as if time travellers were sitting at the table. Combined with the ability to talk face to face across the table, this was hypothesised to generate a dynamic play experience and unique perspectives on fundamental notions of sustainability and leadership.

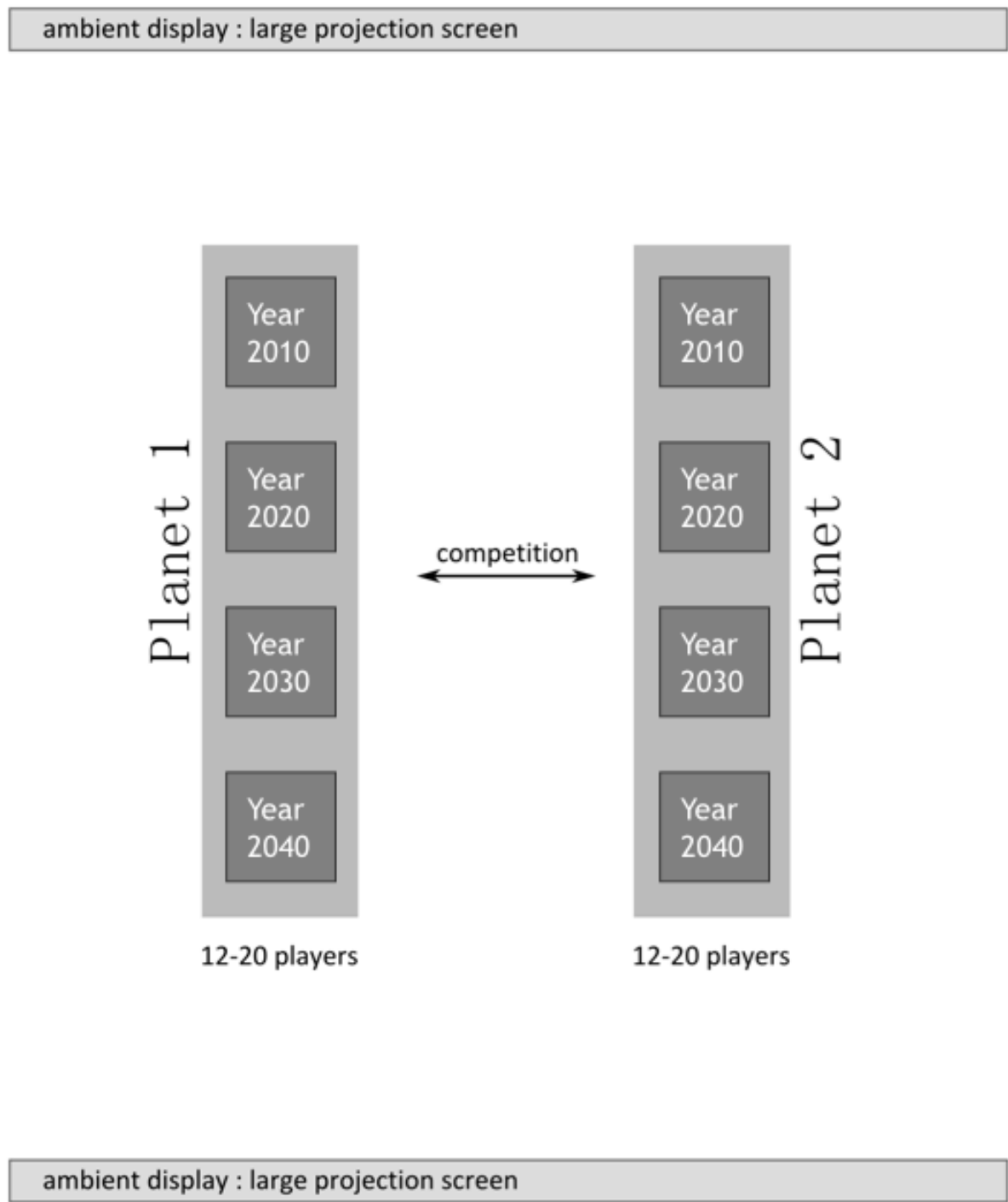


Figure 4.9.: Two competing teams (planets) at large tables

Potential for ambient information

Finally, this sketch made some preliminary attempts at exploring the use of ambient displays. The intention was to provide participants with additional information, such as team-specific feedback, hints, background information or rules of the simulation. The possibility of supplying partial, false or misleading information was also considered, in order to increase the overall difficulty and provoke critical interpretation of given data.

4.7.1.3. Interface design (i3)

Tablets were suggested as input devices, based on the conjecture that a horizontal tablet could be used and shared like a “tiny tabletop” (Rick 2012); and furthermore, tablets might be combined in order to emulate the affordances of a large interactive tabletop. Little academic knowledge had been available regarding how groups can share multiple tablets. Therefore, lab tests were planned to investigate, for instance, whether participants would pick the devices up and pass them around, and whether this would benefit collaboration.

Tablet interface

A simple tablet interface was devised with the aims to be easily shareable. This implied little text, large fonts, simple buttons and no scrolling (see Figure 4.10).

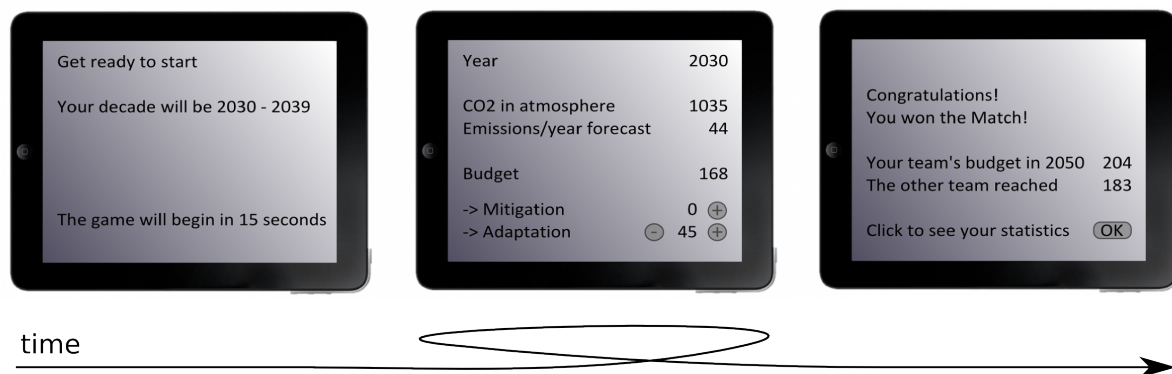


Figure 4.10.: Three states of the tablet interface: Before the simulation (left), during one of several rounds of decision-making (centre) and after the simulation (right).

Ambient information

Two large display were initially suggested (e.g., projectors or flat panel displays), showing exactly identical content. As opposed to using just a single display, the second display was intended to avoid physical crowding and to help all participants see the information easily, regardless of their position in the room. As a first guess, it was considered to have fragments of relevant information appearing and disappearing on the displays.

4.7.2. Evaluation

This sketch was evaluated in two ways: (a) by discussing it informally in our research group and (b) by testing a rough prototype in the lab with colleagues.

The goal of (a) was to elicit further questions regarding each of the Triple-i layers, whereas (b) was to investigate a series of practical and pedagogical questions, most importantly:

- Can a distributed simulation of global climate change be designed for use with multiple input devices?
- Will using the simulation help participants understand the underlying model?

4.7.2.1. Talking about the sketch with other researchers

The sketch triggered many question about the ambient display, including its intended use and benefits, the nature of its content and how this content should be presented. For instance, should information be permanent or temporary? Complete or partial? Cohesive or fragmented? Dynamic or static? Easy or difficult to read and interpret? Graphical, textual or numerical? Should its purpose be to assist, inspire, instruct, distract, mislead, etc.?

Various tentative suggestions were collected. It was unanimously agreed that the overall purpose of the ambient displays should be to present low-level data that is relevant to the groups' problem solving, but that the display should not do any thinking for them, i.e., no hints or other cognitive scaffolding. Moreover, the displays should not distract or mislead, unless later evaluations suggest that this might be desired, e.g., to make the game more challenging. Finally, the displays should be easy to read for all participants, including older people and people with visual impairment.

Besides these rather broad and preliminary assumptions, more detailed decisions about the displays and interfaces depended on better clarity about the concrete problem solving task. Therefore, it was necessary to first specify the concrete climate model and formal game objectives - and to see how users appropriate these in a group setting. In Triple-i terms, this conclusion boiled down to the same old principle of clarifying instructional goals first, then interaction, then interfaces.

4.7.2.2. Prototype testing

A mathematical model of global climate economics was devised, based on some assumptions from the domain literature. The model was implemented as a server-based Java application, allowing several computing devices to connect as clients. The client interface was a simple page with some text and numbers. For lack of available tablet devices, participants were invited to bring their laptops. This was considered sufficient for testing whether people would understand the model. Participants were recruited among computing students and faculty who were interested in the idea and keen to give informed feedback.

An informal test run was conducted in a meeting room with eight participants. Participants were encouraged to talk freely during the activity and the subsequent informal discussion which aimed at highlighting problems of understanding and their likely causes.

Generally, the idea of a distributed climate simulation resonated well among all participants. Surprisingly, however, despite my explanations of how the simulation worked, participants struggled to understand the simulation well enough to make informed decisions.

This was surprising insofar as the mathematics of the climate model per se were not overly complicated. It was safe to assume that all participants had sufficient mathematical skills to understand all the equations (or the source code) easily. The interfaces were trivial to use - no-one found it difficult to enter the decision values.

Rather, the difficulties were unanimously attributed to the “time travel” element which many participants criticised as complicated and distracting (although some found it an exciting gimmick). Indeed, the simulation represented time in an unusual way, letting participants “tap the timeline” at multiple points by manipulating data at constant intervals. On a conceptual level, this seemed easy for participants to understand. However, the challenge of relating the quantitative outcomes of the simulation back to

specific decisions was overwhelming. This prevented the groups from making inferences about the rules of the model. There was a general sense that decisions made on the “earlier” devices (e.g., 2010, 2020) had a greater effect on the outcome than the “later” devices - which was a true observation - but for the most part the groups were unable to figure out the rules or strategies to achieve better results.

Participants had been unfamiliar with the climate model and many argued that the model itself was complex enough to be interesting, and that time travel made the whole thing extremely complicated and difficult to understand.

The ambient display with the appearing and disappearing information was found surprisingly useless. Particularly, it was suggested that the bits of information are difficult to use and share if they keep disappearing.

4.7.3. Discussion

The sketch raised many important questions about the potential of tablets and ambient information.

4.7.3.1. Instruction evaluation (i1)

From an informal evaluation with computing researchers it was concluded that the design had to be greatly simplified in order to make the model more transparent. The fact that computer researchers consistently found the simulation too complicated to use was a bad sign, especially considering that the actual target population was more diverse with regard to mathematical skills. Therefore, the element of time travel was identified as the main source of confusion, acting as an additional - and arguably unnecessary - layer of complication.

4.7.3.2. Interaction evaluation (i2)

The “Parallel futures” sketch suggested two *major hotspots* (the tables) to bundle most of the overall activity and attention. Each of these was in turn composed of four *minor hotspots*, i.e., the tablet devices. This composite structure was intended to enable fluid sharing of information within teams. Unfortunately, the conducted test did not allow a reliable evaluation of this new interaction structure, as participants were barely able to understand the simulation.

The ambient displays could be seen as additional, visual hotspots.

Feedback from participants revealed a critical issue, namely that “earlier” devices had a greater influence on the team’s outcome than “later” devices. This issue - which had indeed been overlooked in the design - clearly violated the overall goal of supporting equitable contribution (Gameplay Objective number 9), by implying more powerful and less powerful positions within a team (despite having equal roles!). For these reasons, it was decided to get rid of the time travel element altogether and find a simpler alternative, based on a single, linear timeline.

4.7.3.3. Interface evaluation (i3)

Conceivably, there may have been some potential for designing (ambient) data visualisations to make the model more transparent. However, this would not have solved the key problem of inequality within teams.

The ambient display had to be redesigned so that information, once presented, would permanently stay visible - preferably in the same place - so that participants could refer to it continuously.

4.7.4. Summary

The feedback from other researchers on the sketch raised some important issues, as summarised below. The decision to recruit lab participants from the computing department paid off in two crucial ways. First, this population had appropriate skills to judge certain characteristics of the system - in this case, an overall lack of transparency and inequality of power within teams. Another, more practical benefit was that I got some useful technical advice on how to implement future prototypes, particularly regarding network performance, security and compatibility with relevant device platforms.

The following conclusions were drawn from this evaluation:

1. Despite participants formally having equal roles, the sketch implied a subtle risk of creating inequality if computational affordances are bound to inequitable parts of the simulation.
2. Time travel is not a good idea in this context.

3. Rather than years, the input devices should represent something more intuitive and equitable, such as equal-size countries on a planet.
4. Ambient information should be helpful, easy to read, low-level and *permanent*.

The following important questions were raised:

- Would participants understand the climate model easily if it was presented in a simpler and more intuitive way?
- Would this require changing the climate model?
- How could the decision-making be distributed equally within a team?
- What design factors about the ambient displays and interfaces could support this distribution?

4.8. Third sketch: “Four equal regions”

The instructional design was changed insofar as an effort was made to present the climate model in a way that was more direct and transparent for participants to use and figure out. With this new focus in mind, the activity was restructured and the mathematical climate model was refined as described below. Those revisions were transferred to the interactive prototype which, furthermore, was improved to include actual tablets (rather than laptops) and a permanent ambient display.

4.8.1. Design

To reduce complication and minimise contrivance, the activity was restructured so that virtual time was the same on all tablets. Four decision rounds are played, advancing in steps of one decade. i.e., round 1 represents 2010, round 2 represents 2020, and so on.

The four tablets now represented regions rather than decades. Each region could make independent decisions in regard to their regional economy. To counter the previous problem of inequality within teams, the regions were defined to have equal size and equal starting conditions (see Figure 4.11). Moreover, no random factors were included in the model. Therefore, regions that always made identical decisions would get identical results.

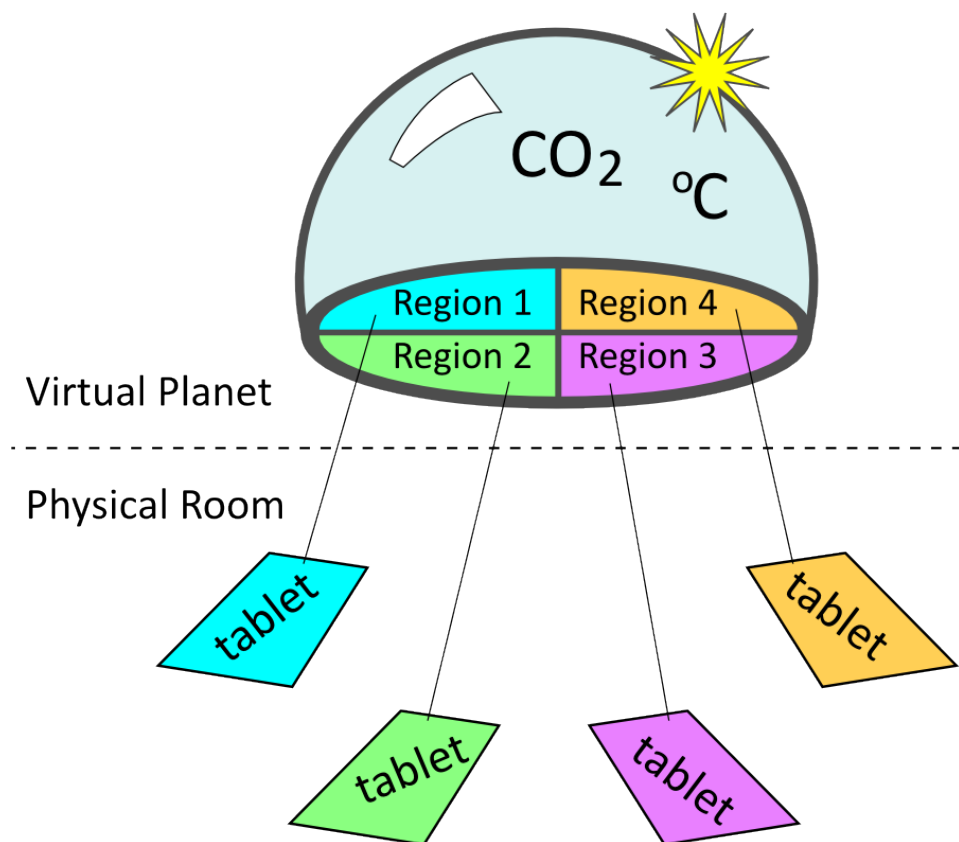


Figure 4.11.: An simplified, Earth-like planet with four regional governments. Each region is controlled using a tablet interface.

4.8.1.1. Instruction design (i1)

A collaborative balancing problem

Each region can decide independently how much it spends on (a) greenhouse gas mitigation and (b) adaptation to climate change. This is a balancing problem, since the one hand, mitigation and adaptation cost money. On the other hand, both also bear a potential to save money indirectly by reducing climate-related damage in different ways. Balancing investments over repeated cycles (decades) becomes a challenge towards the goal of maximising income. The winner of the game was defined as “the planet with the highest global income by 2050”.

Tragedy of the commons

An important property of the climate model was that selfish strategies would fail to produce optimal results, as a common resource was depleted. This principle is commonly known as “Tragedy of commons” (Hardin 1968) and essential to sustainability thinking.

In this model the common “resource” (or variable) was *temperature increase*, which was global across all regions of a planet. The cost and efficiency curves (see below) were designed such that it was eventually better (cheaper) for the whole team if every region contributed equally to mitigation. This was in contrast to an exploitative strategy known as “freeloading”, where one party (region) would let the others do all the work (mitigation spending) whilst benefitting from the shared outcome (low temperature).

4.8.1.2. Devising a climate model

Figure 4.12 shows the key variables in the model and how they influence each other. The round rectangles (blue) represent values that are shown to participants as numbers. The remaining rectangles (grey) represent values that are hidden from participants. The fingers indicate points where participants could control the simulation by entering their spending decisions.

In short, by spending money on mitigation, regions can reduce their amount of CO₂ emissions. The more CO₂ cumulates in the atmosphere, the higher the increase in global temperature and subsequently the expected (baseline) damage to the economy. By

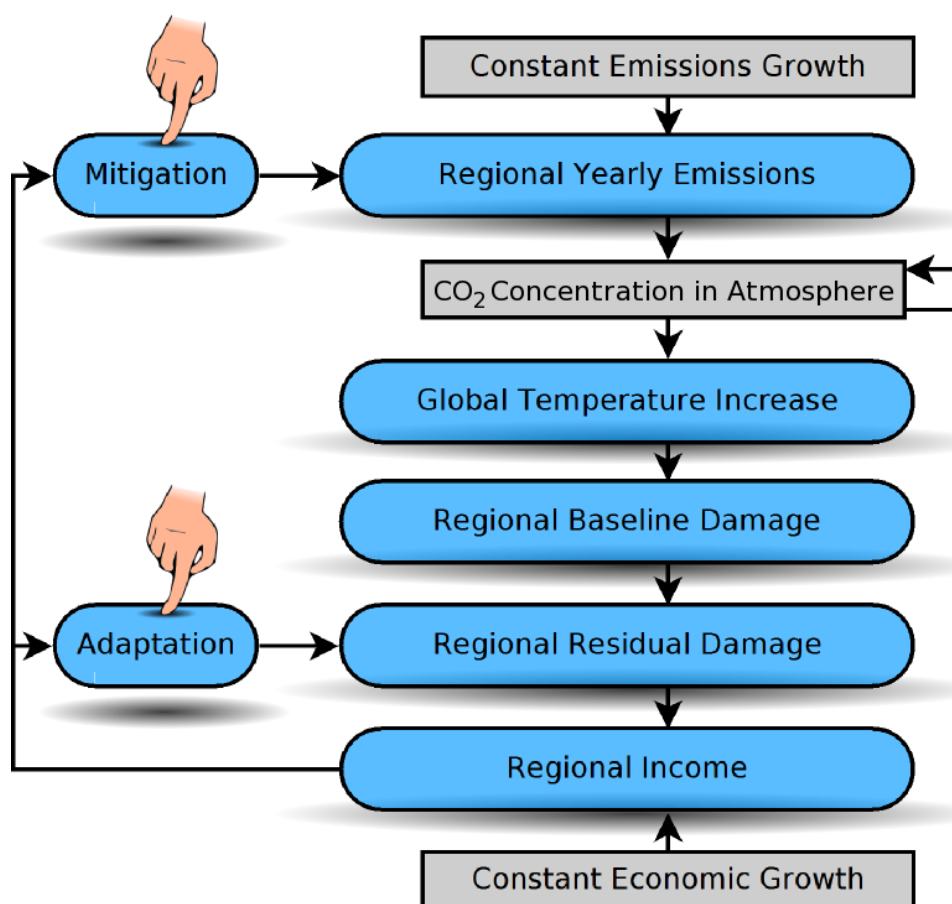


Figure 4.12.: Schematic of the economic model.

spending on adaptation, some of this damage can be reduced. The residual damage for each region is subtracted from that region's income.

The climate model represents a very high-level abstraction of the current mainstream literature on the topic. Knowing how this literature was used in the preceding course lectures, it was safe to assume that participants had a shared understanding of the abstractions used in this model. Without going into too much detail, the most important concepts are explained below for the sake of clarity.

- **CO₂** is the only greenhouse gas considered in this simulation.
- **CO₂ mitigation** in reality may involve a variety of measures ranging from consumer education to policy changes in energy, transport, construction and so on. Those measures vary in cost and efficiency. It is assumed that some of those measures can directly save money, such as automatic light switches, whereas others come at a cost. Consequently, reducing a small amount of CO₂ is assumed to be relatively cheap, whereas reducing large amounts is much more costly, as more of the expensive measures need to be added. To determine how much mitigation can be done at what cost, we used a so-called *marginal abatement cost curve*, aka MAC curve, that was based on the Global Greenhouse Gas Abatement Cost Curve (Enkvist, Dinkel & Lin 2010).
- **Adaptation to climate change** can in reality entail research into drought-resistant crops, building higher sea walls, etc.. In reality, of course, the economic efficiency of any such measure depends on a range of factors, including complex and uncertain factors. Again this model takes only the total spending and assumes average efficiency curves to represent an optimal mix of measures. By spending on adaptation, some of the baseline damage can be reduced (within limits), resulting in a reduced impact on the economy. The Stern Review (Stern 2006) informed some of the numerical assumptions in our model.

4.8.1.3. Implementing the model as an interactive spreadsheet

The climate model was initially built and tested as an interactive Excel spreadsheet. It was then translated into a Java desktop application for further testing and optimisation³.

³The optimised Java version of the model was later integrated into the final application as part of the server-side code

The spreadsheet offered input cells for mitigation and adaptation in each decade between 2010 and 2050 (see Figure 4.13). By typing numbers into those input cells, one could see the outputs change in real-time. This method allowed us to quickly explore different strategies and tweak the model until the outcomes made sense.

| | A | B | C | D | E |
|----|--|--------|--------|--------|--------|
| 1 | | | | | |
| 2 | Round 1 | | | | |
| 3 | | Ipad 1 | Ipad 2 | Ipad 3 | Ipad 4 |
| 4 | Now | | | | |
| 5 | Year | 2010 | 2010 | 2010 | 2010 |
| 6 | Regional GNI | 17.5 | 17.5 | 17.5 | 17.5 |
| 7 | Regional CO2 | 7.5 | 7.5 | 7.5 | 7.5 |
| 8 | Mitigation Decision (GtC) | 1 | 0 | 2 | 5 |
| 9 | Adaptation Decision (\$100 billion) | 3 | 3 | 1 | 0 |
| 10 | Forecast for 2020 based on your decisions this round | | | | |
| 11 | Temp increase from 2010 (oC) | 1.254 | 1.254 | 1.254 | 1.254 |
| 12 | Global cumulative additional CO2 from 2010 (GtCO2) | 314 | 314 | 314 | 314 |
| 13 | Your Regional CO2 in 2020 | 7.8 | 8.8 | 6.8 | 3.8 |
| 14 | Your GNI in 2020 | 23.2 | 23.2 | 22.6 | 21.5 |
| 15 | | | | | |
| 16 | | | | | |
| 17 | Round 2 | | | | |
| 18 | | Ipad 1 | Ipad 2 | Ipad 3 | Ipad 4 |
| 19 | Now | | | | |
| 20 | Year | 2020 | 2020 | 2020 | 2020 |
| 21 | Regional GNI | 23.2 | 23.2 | 22.6 | 21.5 |
| 22 | Regional CO2 | 7.8 | 8.8 | 6.8 | 3.8 |
| 23 | Mitigation Decision (GtC) | 5 | 5 | 5 | 5 |
| 24 | Adaptation Decision (\$100 billion) | 0 | 0 | 0 | 0 |
| 25 | Forecast for 2030 based on your decisions this round | | | | |

Figure 4.13.: Frontend of the interactive climate model spreadsheet.

In the figure, the columns B, C, D and E represent the four tablets, i.e., the four regions on the planet. The black number cells represent input fields. For example, the number 2 entered in cell D8 means that region 3 chose to mitigate 2 gigatonnes of CO₂ in the first

Documenting the assumptions and equations was done using multiple worksheet layers and integrated comments. This allowed the whole team to understand the rationale behind each concept and each equation. Figure 4.14 shows an excerpt of this documentation.

Figure 4.14.: Fully commented backend worksheet, annotated with literature references and explanations of domain-specific concepts and equations.

Learning from the previous mistakes, this sketch emphasised equality of power within each team. For this purpose, the regions were designed to be identical from the start and subject to identical rules. This was, of course, an enormous simplification compared to real countries. This simplification shows how fidelity was sacrificed for the greater goal of equitable contribution between participants (Gameplay Objective number 9).

4.8.1.5. Interface design (i3)

To make the revised climate model accessible in a group setting, it was re-implemented in Java (using the equations and comments in the spreadsheet as a template) and integrated in the interactive prototype.

Ambient information

The ambient display was re-conceptualised from scratch. Whereas the previous version included fragments of information that appeared and disappearing at incongruous times and positions, the new design emphasised that data should be presented in a way that is well-structured, cohesive and permanent. The simple organising principle was that after each round of decision making, a new line of data would be added to the bottom of a table. Figure 4.15 shows an initial spreadsheet-based design. Figure 4.16 shows how this was implemented in the interactive prototype.

| | | | | | | | |
|-----------|-------------|------------|------------|---------------------|---------|---------------------|---------------------|
| Planet A | | | | | | | |
| Match 1 | | | | | | | |
| Round | GNI | Mitigation | Adaptation | Emissions | TempInc | BaseDmg | ResDmg |
| 1 | 18/18/18/18 | 2/0/0/1 | 3/4/4/2 | 8.0/8.0/8.0/8.0 | 1.3 | 1.9/1.9/1.9/1.9 | 1.5/1.7/1.7/1.6 |
| 2 | 21/18/22/20 | 3/0/3/3 | 4/5/5/2 | 11.3/14.2/15.3/13.3 | 2.3 | 11.3/14.2/15.3/13.3 | 11.3/14.2/15.3/13.3 |
| 3 | 33/32/29/34 | 4/3/2/3 | 4/3/4/5 | 13.2/13.2/15.6/12.8 | 3.7 | 13.2/13.2/15.6/12.8 | 13.2/13.2/15.6/12.8 |
| 4 | 35/37/35/36 | 4/2/2/4 | 5/6/5/7 | 13.2/13.2/15.6/12.8 | 4.4 | 13.2/13.2/15.6/12.8 | 13.2/13.2/15.6/12.8 |
| Final GWI | 132 | | | | | | |

Figure 4.15.: Four decision rounds on the ambient display. GNI stands for “gross national income” and GWI for “gross world income”. The numbers are mock data.

PLANET A

4Decades

MATCH 1

| Round | GNI | | | | Mitig | Adapt | Emissions | | | | TempInc | BaseDmg | | | | ResiDmg | | | |
|------------------|------|------|------|------|-------|-------|-----------|------|------|------|---------|---------|-----|-----|-----|---------|-----|-----|--|
| 1 | 17.5 | 17.5 | 17.5 | 17.5 | 0.0 | 0.0 | 7.5 | 7.5 | 7.5 | 7.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 2 | 22.2 | 22.2 | 22.2 | 22.2 | 0.0 | 0.0 | 8.7 | 8.7 | 8.7 | 8.7 | 1.4 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | |
| 3 | 25.5 | 25.5 | 25.5 | 25.5 | 0.0 | 0.0 | 10.0 | 10.0 | 10.0 | 10.0 | 2.9 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | |
| 4 | 26.0 | 26.0 | 26.0 | 26.0 | 0.0 | 0.0 | 11.2 | 11.2 | 11.2 | 11.2 | 4.6 | 9.4 | 9.4 | 9.4 | 9.4 | 9.4 | 9.4 | 9.4 | |
| Final GWI = 89.9 | | | | | | | | | | | | | | | | | | | |

Figure 4.16.: The equivalent in the interactive prototype. The example numbers illustrate what happens when no spendings are made: High temperature increase and high residual damage.

Tablet interface

The tablet interface was revised, with the aim to expose all the essential variables that participants needed to see in order to make strategic decisions. (Figure 4.17). In order to support decision making, the variables were ordered from top to bottom, following the flow of data in the model. I.e., the top section shows the given variables: year, regional income and emissions. The middle section contains the variables that participants can increase and decrease using large, easy to use buttons. The bottom section shows a forecast of predicted outcomes for the next decade. The numbers in the forecast area respond in real time to decisions being changed, not just on this region but also on the other regions (tablets) on the same planet (team). This was hypothesised to encourage participants to pay attention to their teammates' decisions.

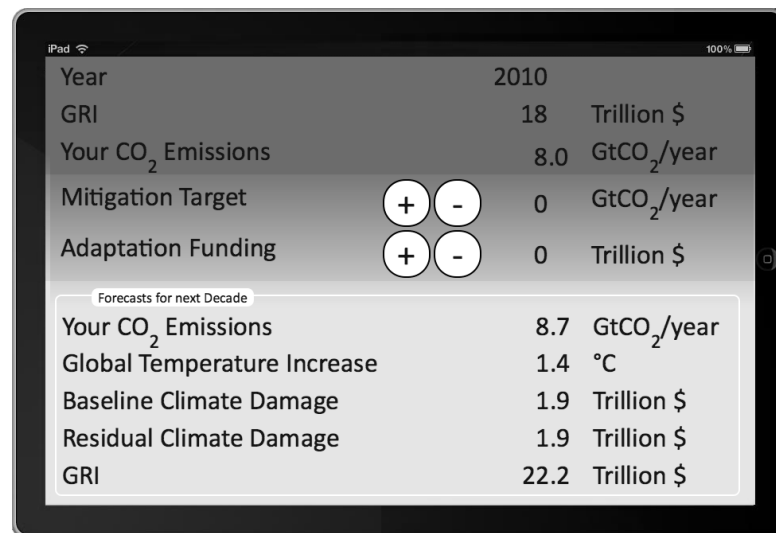


Figure 4.17.: The tablet interface during decision making.

4.8.2. Evaluation

The evaluation of this version was guided by two questions:

1. Will the model consistently produce plausible outputs for plausible inputs? By 'plausible' we mean that an expert facilitator could explain the numbers to participants conclusively with reference to the theory taught in the course lectures.
2. Is the new version of the simulation easy enough for participants to understand and use, so that they can make meaningful decisions in the game?

The first question was addressed (a) by asking domain experts for their opinion on the spreadsheet and (b) testing the model against large amounts of random data to detect any errors in the equations. The second question was addressed in another lab test.

4.8.2.1. Expert feedback

Once the spreadsheet was complete, Stephen consulted several modelling experts and colleagues who teach climate and economics at university level. He demonstrated the spreadsheet, explained the rationale and asked their opinions on whether the model would be suitable for teaching. Those conversations confirmed that the assumptions and outputs of the model were all within a plausible range, although some of them would need explanation. It was therefore advised that a competent facilitator should be there to prevent erroneous conclusions.

Figure 4.18 shows two example scenarios with and without spending on adaptation.

4.8.2.2. Numerical testing

Once the model appeared fairly mature from a pedagogical perspective, it underwent further computational tests. For this purpose the model was translated into Java. Large amounts of random data were fed into the model, aiming to detect combinations of inputs that would produce highly implausible outcomes. A second algorithm verified that the space of possible strategies did indeed represent a 'tragedy of the commons' without any loopholes.

A number of optimal strategies were detected for achieving certain goals within the simulation, e.g., maximising global income by 2050 or keeping global warming below 2 degrees. We critically examined the plausibility of these strategies in the context of the current climate literature. This stage of evaluation led to minor adaptations in the model, such as adjusting the efficiency of adaptation. After a few repetitions of tweaking and testing, the climate model was considered mature for deployment in an expert-facilitated setting.

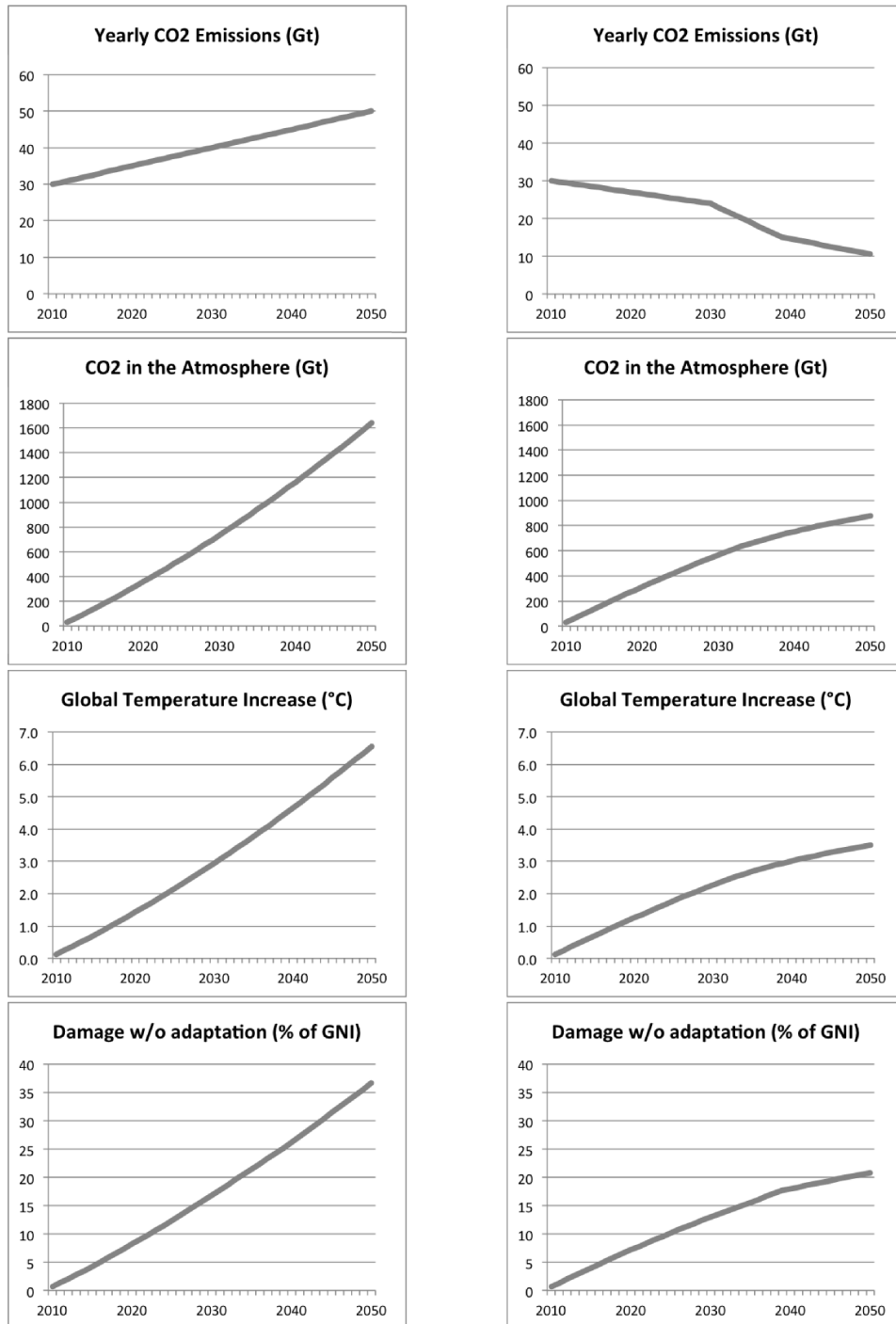


Figure 4.18.: Example outcomes that the model predicts. Left column: Without spending on mitigation. Right column: With a modest and continuous spending on mitigation. Note the higher damage (cost) in the no-spending case.

4.8.2.3. Lab testing

Another lab trial was conducted to test whether the design revisions were effective, i.e., whether the new and simplified version was usable in a group setting. Whereas the previous test had failed due to an overly complicated design and a flawed ambient display, it was hypothesised that the revised version would allow participants (a) to develop a functional understanding of the simulation, (b) to make meaningful decisions and (c) to use the ambient display for reflecting on outcomes. Furthermore, the evaluation aimed to highlight feasible ways of sharing tablets and ambient information in a large group.

11 volunteers took part, of which 7 worked in computing. Participants received not monetary compensation for taking part but free snacks and drinks were provided. The session began with a short introductory lecture on the topic. Two matches were played, followed by a 20 minute joint debriefing discussion. Video, audio and photos were taken. To explore how participants appropriated the distributed simulation, no instructions were given regarding where to move or how to share the devices.

Findings regarding instruction (i1)

- **Finding 1: The model was understood sufficiently.**

Despite having little previous information about the simulation, every participant who had grabbed a tablet immediately started playing with the changing numbers by pressing the four buttons, as soon as the simulation began.

Over the course of two matches (roughly 10 minutes per match) the groups had figured out the essential mechanisms enough to make strategic decisions. This was clearly evidenced in the debriefing. For instance, one participant pointed out that the common resource enabled a freeloading strategy:

“Well, it’s just interesting, when we actually were all together in the corner here and actually one group could just sit doing absolutely nothing and watch their profits go up because somebody else had actually changed something. And the whole planet would benefit - but actually you as a region did nothing.” (P7)

Another participant noted that the best outcomes could be achieved if all regions collaborated:

“At that point everyone just agrees to press the same button. And do the same thing. Because they all agree that that’s the optimal strategy for the planet.” (P3)

- **Finding 2: Improvements could be made to the delivery.**

Despite a generally good overall understanding, some aspects were unclear to some participants, especially in the beginning. Participants attributed these difficulties to two factors: Firstly, the fact that some of the model variables were not displayed, for instance:

“This notion of interest, was there actually interest being grown on your gross national income? It was a bit unclear because you couldn’t see.” (P3)

Secondly, the introduction would be clearer if certain aspects were emphasised or deemphasised, according to multiple suggestions from the group. For instance, one participant requested clarification about whether or not the forecasts were accurate:

“There is the forecast, erm, temperature change. Was that actually what resulted at the end of the ten years - or where there some surprises?” (P6)

Conversely, some others aspects could be deemphasised in the introduction, such as the rate of economic growth:

“I was misled by the 3.3% compound interest thing. Because that’s kind of more insignificant compared to potential benefits from, erm, realising savings through adaptation.” (P6)

Finally, generic suggestions were made regarding the mode of delivery. These included the idea to explain the rules in a more visual way, by showing a diagram of the economic model, a powerpoint or a video. Moreover, it was suggested to label the tablets (e.g., with colours) to make it easier to see who is who.

Group interaction (i2) using the tablets

- **Finding 3: The tablets were controlled by individuals and pairs, either on the table or held while standing upright (Figure 4.19).**



Figure 4.19.: All participants were standing.

- **Finding 4: Handheld use made sharing in groups difficult.**

In the case of handheld use, a recurring observation was that participants tilted the tablets in order to show the screen to others. However, tilting seemed to somewhat conflict with the need to enter values (Figure 4.20). Although the quality of the screens was essentially suitable for reading from an angle, persons holding an iPad would rarely stand perfectly still, thus making it difficult for others to read the text.



Figure 4.20.: Holding the tablet enabled easy sharing or entering values - but not simultaneously.

Group interaction (i2) using the ambient display

- **Finding 5: Participants were able to interpret the ambient information.**

The ambient display allowed participants to relate their decisions to outcomes and hypothesise about the model.

- **Finding 6: Small text on the ambient display required some participants to stand close, thus fragmenting the group.**

In order to read the numbers, participants had to stand fairly close to the display (Figure 4.21). In this situation participants were focused on the display, without using the tablets or talking to teammates who were using tablets.

Use of tablet interfaces (i3).

- **Finding 7: Individuals were immediately engaged in playing with the numbers using the tablets.**
- **Finding 8: Using the forecast, participants quickly realised that the real-time forecast numbers were influenced by their own actions as well as their teammates' actions.**



Figure 4.21.: Pairs studying the ambient information. Left: after the first match; Right: after the second match.

The observations indicated that the interfaces were basically usable. No difficulties were raised explicitly, except for one comment regarding the naming of domain concepts:

“I got confused between GNI and yearly budget.” (P6)

Other participants agreed that the naming should be consistent (a) between the introduction and the simulation and (b) between the different devices and displays. By accurately pointing out such minor deficiencies, participants demonstrated a pragmatic understanding of the simulation and contributed useful feedback towards improving the design.

4.8.3. Discussion and changes

The lab test revealed useful insights that informed relevant changes to the prototype and the delivery. These changes pertained to all Triple-i layers, as detailed below.

4.8.3.1. Changes to instruction design (i1)

Considering that the participants in this test had been given only a brief introduction to the topic and the key variables in the model - unlike the real target group - the group understood the model surprisingly quickly and accurately. In contrast to the previous version that involved a complicated construct of time travel, the familiar concept of economic and political regions on a planet was intuitively graspable.

The teacher found the session useful for rehearsing his introduction and debriefing of the activity. Participants' comments and questions allowed him to refine the initial explanations to make them clear, exhaustive and concise. The teacher emphasised that

the goal was to make the rules absolutely clear, so that the learning could focus on more high-level aspects, such as strategies, rather than figuring out the rules.

Rather than showing a video or powerpoint slides, it was decided to do a short practice round (that did not count towards winning) and a quick walk through each element of the interface.

4.8.3.2. Problems with interaction (i2): Orientation, crowding, awareness

Problems with sharing were observed when participants held the tablets in their hands. Therefore, it was decided to encourage users to leave the tablets resting on a horizontal surface.

To achieve this, initial ideas considered mounting the tablets on stands at chest height (e.g. cocktail tables). This would allow participants to stand comfortably while using the tablets. Moreover, participants would have much freedom to walk opportunistically between the tablets (hotspots) and approach the tablets from any direction. However, there were at least three problems with this option:

1. **orientation.** If groups stood around a tablet in a circle, then some of them would have to read upside down, compromising the goal of equal access.
2. **crowding.** Only a small number of people (about three or four) would be able to stand close enough to the tablet to use the touch screen.
3. **awareness.** If a tablet was surrounded by group, people outside the group would be unable to see the screen.

To achieve a feasible compromise with regard to the above factors, an alternative solution was considered. This involved one large table for each team, with the tablets facing outward.

4.8.3.3. Proposed solution for interaction (i2): The Panoptcake layout

Figure 4.22 illustrates this idea which aims to combine some of the benefits of Foucault's Panopticon with those of cake. Notable aspects of the Panoptcake layout are:

- Participants of the same team are arranged in a circle around a table. Facing inwards allows them to easily see each other and talk across the table.
- The table itself should be approximately circular and divided - like a cake - into equal sectors.
- Each sector contains a tablet positioned in front of two sitting participants, as if these two participants were sharing a plate of cake. They are called the "players".
- The other team members - called the "analysts" - distribute themselves around the table, where they can easily read the tablets over the players' shoulders, listen to the discussions and take part verbally.
- Since the analysts are standing and able to walk, they are in a privileged position for using the ambient displays.
- Analysts and players are encouraged to swap roles between repeated cycles of the simulation.
- Unlike Foucault's proposal, participation is voluntary and there is no central observer.

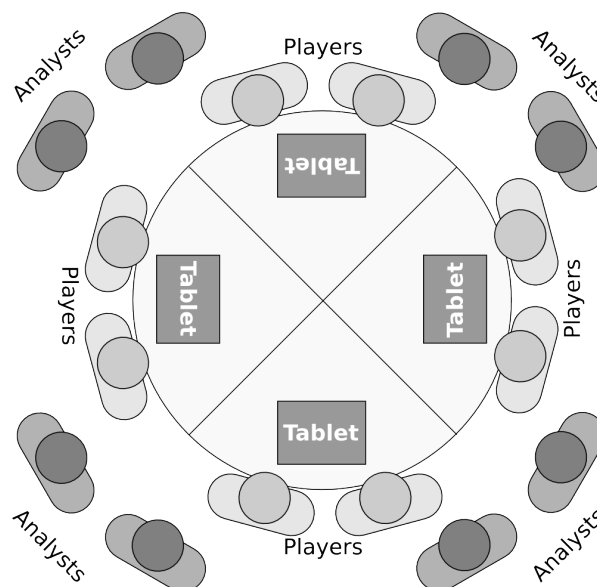


Figure 4.22.: The Panoptcake layout

This layout was partly motivated by the conjecture that social protocol would prevent participants from snatching a tablet away from a pair sitting in front of it. At least in a typical cake situation this would be considered rude. Moreover, it was hypothesised that polite individuals, having received early training in equitable cake-sharing, would intuitively transfer this cultural meme to the interactive devices. Therefore, it was presumed that the target group would find negotiating control at the tablets interfaces easy.

Providing chairs for the players was considered promising in this context. The downside of players' reduced mobility seemed outweighed by the following reasons:

- to avoid physical strain from bending over the table;
- to reinforce a polite and laid-back, coffee party atmosphere - rather than a standing buffet situation where people tend to get in each other's way;
- to free the "upper layer" (head and shoulder height) for the analysts to look more easily over the shoulders and around the room.

Applying the Panoptcake layout in practice

Given that circular tables in the suggested format were not easily available, the layout was varied to allow for square-shaped tables instead. These are usually easy to approximate by moving standard, rectangular tables together. Moreover, one side of the table was reserved for the team's ambient display projector. This led to a revised room layout (Figure 4.23) with some further specifications that the course organisers could use in search for a suitable room at the conference centre.

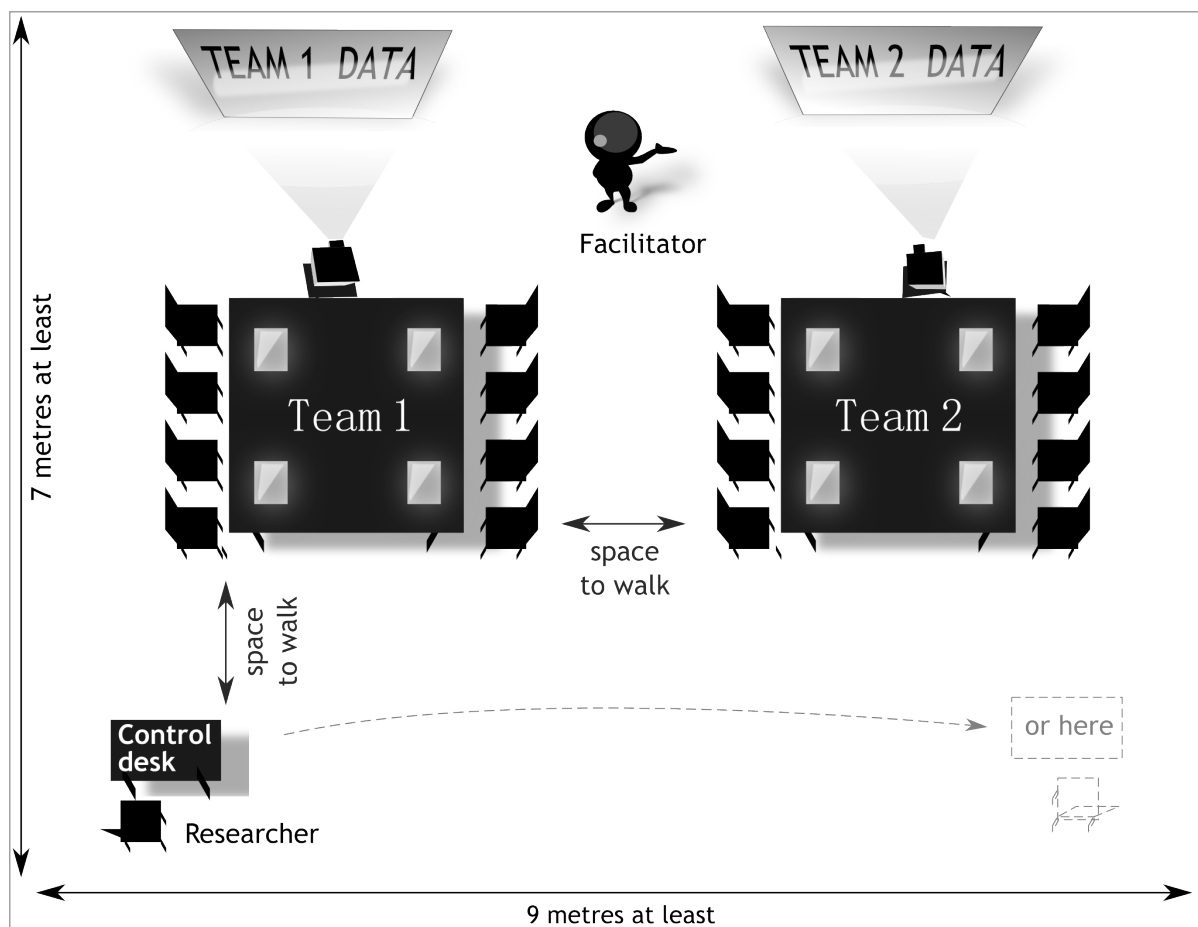


Figure 4.23.: A revised room layout, suggesting the facilitator close to the displays and a researcher (myself) controlling the simulation from a far corner of the room.

4.8.3.4. Changes to the interface design (i3)

Tablet interfaces:

No changes to the tablet interface were found necessary. The fact that individuals were immediately able to play with the numbers using the tablets was strong evidence for the interface's basic usability.

When the tablets were on the table, pairs could share them easily. There appeared to be a tendency for participants to take their hands off the screen after pressing a button. This may have been simply a matter of polite behaviour or partly related to the fact that some numbers were changing just below the buttons and people avoided occluding the information, consciously or not. Surprisingly, participants had just enough time to remove their hand before the numbers changed in the forecast. This was due to an unintended delay in the network connection, to do with the tablets polling information from the server via TCP/IP at regular intervals. Although it would have been possible to minimise this delay through some optimisation, it was decided to leave it in as it appeared to benefit the shared user experience.

Detailed in-situ observations were required at this stage to investigate further aspects of shareability under real conditions, and to elicit any unexpected factors.

Ambient information:

For using the ambient information in the lab test, people had to stand close to the display. Larger screens (projectors) were suggested in order to resolve this limitation and allow participants to read the information at a glance from any position, including the people using the tablets. This kind of equal accessibility was required by Gameplay Objective 9.

For the same purpose, the visual design of the ambient display was improved to further increase readability from a distance. Most notably, as seen in Figure 4.24, the text was enlarged compared to the old version (Figure 4.16). This was achieved, among other things, by using smaller margins and getting rid of the "Round" column. The column for "GNI" (Gross National Income) was renamed "GRI" (Gross Regional Income) for naming consistency and moved to the right, in order to have decisions and results arranged in reading order from left to right. To save vertical space, the "Final GWI" (Gross World Income) label was moved up to the same height as "Match 1", "Match 2", respectively. Finally, grey backgrounds with gradients were introduced to visually distinguish clusters of numbers. This made it possible to move the clusters closer together, thus creating space for larger text.

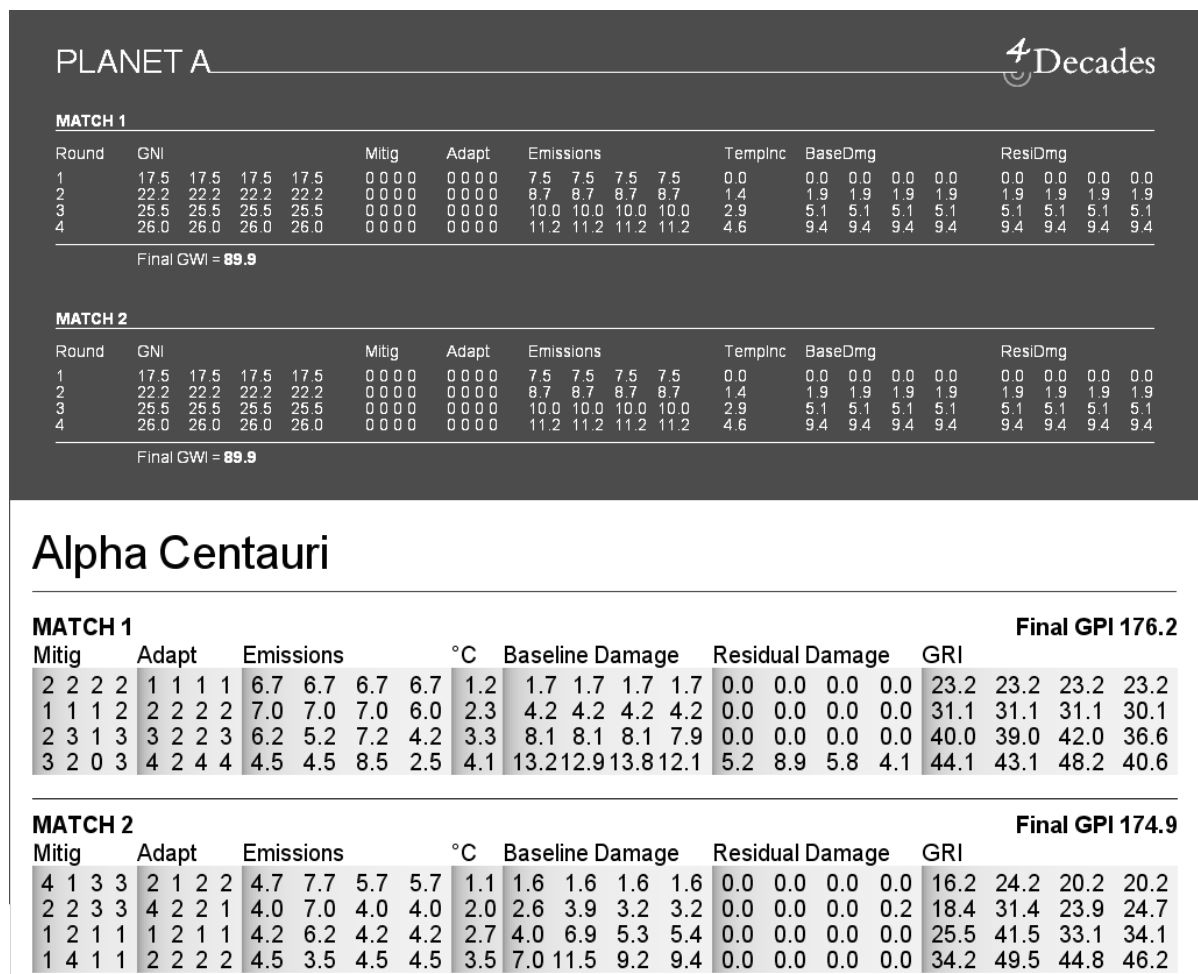


Figure 4.24.: The ambient display in an early stage before the final lab test (top half), compared to the first in-situ evaluation (bottom half). Each showing 2 out of 4 matches.

Besides, the team names were changed. In response to a suggestion from the course committee arguing that “Planet A” and “Planet B” sounded a bit dry, the planets were given tongue-in-cheek names: “Alpha Centauri” and “Betelgeuse”.

4.8.4. Summary

After consultation with climate experts and algorithmic testing, the model was considered sufficiently transparent and plausible for use in a facilitator-led group discussion. The facilitator’s responsibilities would include answering participants’ questions about the model when required. The model was based almost exclusively on material taught in the lectures, with only a few minor assumptions added that were not part of the literature.

In summary, the findings supported the goal for groups to be able to understand the model, make decisions and interpret the data on the ambient display. Subject to some minor corrections, the simulation was considered mature for deployment in the wild. These minor corrections included:

- consistent naming and streamlining the introduction (i1);
- the decision to provide chairs to encourage easy sharing of the tablets (i2);
- optimising the ambient display to make the text as large as possible (i3).

The following chapters will describe the setting, the technological implementation of the application and evaluation in real training courses.

4.9. Overall discussion of the lab-based design

Over the course of three sketches and several prototype evaluations, many lessons were learned. Ethnographically informed Gameplay Objectives, approached via the Triple-i hierarchy, guided three consecutive sketches and lab-based evaluations.

4.9.1. Reflections on iterative design and evaluation

The sketches and lab-based user tests allowed me to eliminate a number of solutions that were unlikely to work well in the wild, such as letting participants carry tablets around or presenting the simulation in overly complicated ways (e.g., via a convoluted

time travel system). Each lab test furthermore raised a number of relevant questions and indicated a range of possible alternatives to inform further design.

Although several cycles of iteration and redesign were required, the design nevertheless resulted in a deliverable prototype that was finished on time. This success was partly due to the design process essentially staying focused on the key instructional goals without getting too distracted by the occasional desire to build a pretty interface, stunning visualisations or try out whacky game patterns.

4.9.1.1. Achieving focus through focused evaluation

Arguably, the Triple-i hierarchy played a crucial role in keeping the design focused. This was less a matter of getting things on first try - in fact, the first two sketches failed to meet some essential Gameplay Objectives. Rather, each time an evaluation shattered an idea to bits, adhering to the Triple-i hierarchy allowed me to pick up the pieces in a useful order. Particularly, the principle of prioritising instructional goals over interaction and interface aspects was extremely useful in deciding which aspects of a design were worth keeping, which ones needed to change and which ones needed to go.

4.9.1.2. Reflections on whole-activity lab testing

The lab tests emphasised realistic conditions, including an actual teaching procedure with introduction, simulation and debriefing. Participants were observed and encouraged to give feedback. The focus was on the whole activity, rather than isolated factors.

Benefits

Testing the activity as a whole was found beneficial for several reasons:

1. It allowed the teacher to rehearse the introduction, facilitation and debriefing under conditions that were close to real, from an orchestration point of view.
2. It allowed the researcher to verify that a group could learn to use the interfaces (Gameplay Objective 2), and that they could do so without the need for a teacher's assistance. The latter aspect relates to Gameplay Objective 3, requiring that *one* teacher should be able to handle all the facilitation.
3. Many useful suggestions were contributed by participants.

4. It saved time, compared to testing individual design aspects or participants one by one. Relevant individual experiences, such as initial difficulties interpreting certain numbers, were captured by the test, as participants were able to clearly articulate those experiences in the debriefing.

In this way, a variety of relevant issues pertaining to instruction (i1), interaction (i2) and interface (i3) were successfully elicited and addressed - enough to devise the right changes to move the design forwards in appropriate steps.

Limitations

A general limitation of lab studies is that they may not reliably produce the entire range of salient factors that might emerge in the wild. Rogers et al. (2013) warn that “*a lab study investigates factors as planned by the experimenter, but in-the-wild studies can reveal unexpected effects of the context of use not anticipated in a lab study.*”

To counteract this risk, an effort was made to replicate as much of the *context of use* as possible, such as using the same technology, the same teacher, a similar population (highly educated and polite adults) and a similar procedure. Nevertheless, some important factors could not be replicated, including participants’ level of background knowledge, learning objectives, professional relations and subtleties of social protocol in the course setting. Therefore, a residual risk of unexpected issues was inevitable.

4.9.1.3. Supporting mechanisms for collaboration

When evaluating sketch 1, initial discussions with colleagues had already indicated some shortcomings in the interaction layer (i2). Objections were raised about teams being isolated. However, moving from these broad objections towards concrete design decisions required a more specific diagnosis. This was achieved by applying the Mechanisms for Collaboration framework (Yuill & Rogers 2012) at two levels of group size: (a) within teams and (b) across teams. The result was a radical shift of focus from issues of small-group collaboration to the whole-classroom big picture, leading to an improved design. This result is remarkable for several reasons:

1. Although the framework was originally developed for analysing small-group collaboration, the current example highlights the framework’s unexpected potential as a tool for analysing multiple (simultaneous) scales of collaboration, including large-scale collaboration and collaboration between groups.

2. Applying the framework to a mere sketch (rather than conducting time-consuming user tests) allowed the design to move ahead quickly.

4.9.2. Ecology of devices

The final lab test (sketch 3) revealed some generally useful observations into how large groups can share tablets. Relevant sharing patterns and obstacles were identified. These findings informed the Panoptcake group layout, proposing that two people could easily share a tablet as *drivers* (sitting) whilst being advised by two or more *navigators* (standing). This assumption supported the originally proposed number of 8 tablets in total - subject to further testing in the wild.

The findings suggest that ambient displays can meaningfully complement tablet interfaces. This seems true in terms of physical access as well as problem solving.

4.9.2.1. Physical access

In terms of physical access, participants could easily switch between *looking down* at the tablets and *looking up* at the wall displays. Moreover, while sitting was considered to benefit using the tablets, standing appeared more appropriate for reading the wall displays, since standing participants could easily turn or walk to the displays.

4.9.2.2. Collaborative problem solving

Tablets and ambient displays afforded complementary support for problem solving in the large group, via different mechanisms for collaboration (Figure 4.1). The tablets allowed participants to control the interface. The ambient displays provided background information. Both encouraged mutual awareness, albeit at multiple levels: Whereas the tablets' realtime forecast stimulated awareness *within each team*, the ambient displays' permanent record enabled awareness *between teams*.

4.9.2.3. Complementary scopes in space and time

Figure 4.25 illustrates in abstract terms how the tablets emphasise *instant feedback* on a local scale (left), whereas the ambient displays provide a *permanent record* on a global

| Affordance | Mechanism |
|------------------------------|--|
| Tablets | Control over the interface |
| Ambient displays | Availability of background information |
| Tablets and Ambient displays | Mutual awareness of actions and intentions |

Table 4.1.: Mechanisms for collaboration (Yuill & Rogers 2012) mapped to affordances

scale (right). Arguably, this distinction between local and global is more descriptive than private versus public - primarily because the tablets are not really private. Rather, the proposed design is based on the premise that anyone standing near a tablet has access. Even if a person is not close enough to touch, they may still be close enough to read. Spatial proximity is the salient factor in this regard, rather than personal roles.

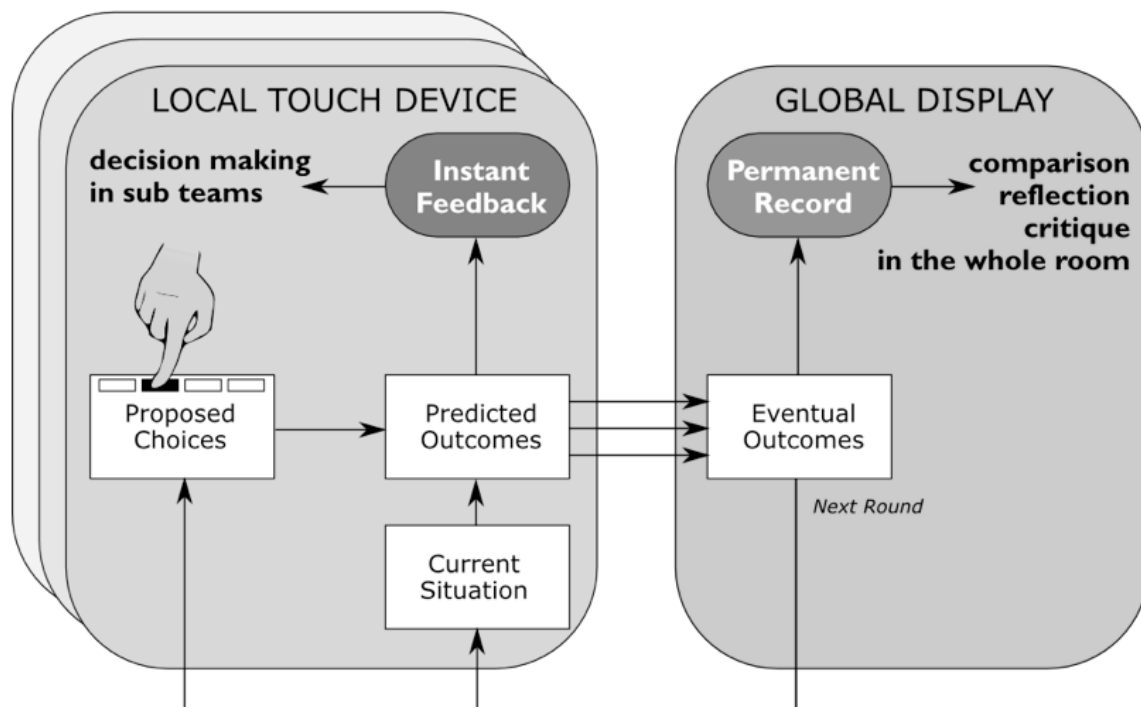


Figure 4.25.: Complementary scopes of a team's tablets (left) and ambient displays (right)

4.9.3. Suggested use qualities for ambient displays

Moreover, the findings suggested that the ambient displays should be designed towards the following use qualities:

1. **Persistent:** Once information has appeared on the screen, it should stay in the same place, allowing opportunistic and repeated use.
2. **Complete:** All the data necessary for informed decision making should be represented.
3. **Cohesive:** Elements should be visually grouped by relation. For instance, one area for decisions and one for outcomes.
4. **Numerical:** Unlike graphical shapes, numbers are generally easy to verbalise, as one can simply read them out loud. Appropriate units and precision should be chosen to balance accuracy and readability. To save display space, units can be hidden if they are clearly defined elsewhere, e.g., on tablets or paper handouts.
5. **Historical:** The data should represent a history of past decisions and outcomes, rather than future predictions or advice.
6. **Accurate:** Participants should be able to rely on the data being a truthful representation of the simulation's state.
7. **Easy to use:** The display should avoid introducing any visual challenges, in order to support equal access and to keep the focus on the task challenge.

4.9.4. Ready for the wild?

Although the lab tests produced relevant insights and a robust prototype, there were some limitations. On the one hand, the lab tests were essential in eliminating a number of design aspects that seemed unlikely to work well in the wild (e.g., time travel, carrying tablets around). They also revealed a number of promising alternatives. On the other hand, due to a lack of “real” participants in the real situation, the lab tests offered limited assurance that the current design was indeed feasible under realistic conditions.

Therefore, 7 out of 9 Gameplay Objectives were still pending to be evaluated in the wild before reliable conclusions could be drawn (Table 4.2).

| N | Gameplay Objective | status |
|---|---|--------------|
| 1 | Allow participants to apply relevant course knowledge in practice through decision-making and debate | achieved |
| 2 | Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use | achieved |
| 3 | Allow a single teacher to facilitate and debrief the activity, using an ecology of displays | to be tested |
| 4 | Encourage participants to reflect on assumptions through decision making and discussion | to be tested |
| 5 | Allow participants to refresh from exhausting lectures by moving and talking | to be tested |
| 6 | Encourage negotiation by letting participants follow different goals but requiring the consent of teammates | to be tested |
| 7 | Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play | to be tested |
| 8 | Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data | to be tested |
| 9 | Allow all participants to contribute equitably to the activity and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics | to be tested |

Table 4.2.: Gameplay Objectives

4.10. Chapter summary and conclusion

This chapter reported on the ethnographical preparation and lab-based design of the 4Decades simulation. Based on 9 concrete Gameplay Objectives and guided by the Triple-i hierarchy, three sketches were used to explore solutions for making large-group collaboration feasible through computer-supported role-play. The evaluation included two interactive prototypes that were tested with large groups of participants in the lab.

Although generalising from the lab-based findings to real educational settings would be premature - due to the inherent limitations of the lab situation - the findings were instrumental in informing design iterations. The fact that these iterations rapidly led to a usable prototype application provides validation for the chosen iterative design strategy. Specifically, the Triple-i hierarchy was found useful for organising the ideation, specification, evaluation and discussion of design iterations. Moreover, support was found for the Mechanisms for Collaboration framework (Yuill & Rogers 2012) being applicable in a large group context.

All 9 Gameplay Objectives were addressed and verified to the extent that the lab situation allowed. The next chapter will report on a series of in-the-wild evaluations, using these objectives to guide the analysis.

Chapter 5.

Analysis 1: In-situ feasibility of the 4Decades simulation

The following three chapters describe how the 4Decades simulation was evaluated in professional training. Each chapter focuses on a different part of the evaluation. I refer to these three parts as Analysis 1, Analysis 2 and Analysis 3.

This chapter describes Analysis 1 (Section 5.2) after a brief overview of the three analyses. The overview is there to clarify the relations between the analyses and to summarise the overall research aims, method, setting and technical implementation.

5.1. Overview of Chapters 5-7

- Chapter 5 (Analysis 1) verifies the basic feasibility of the simulation in a training course, based on a single session that lasted one hour.
- Chapter 6 (Analysis 2) describes a design iteration that additionally allowed participants to change the rules of the simulation. Collaborative rule-changing is discussed with regard to pedagogical benefits and practical feasibility.
- Chapter 7 (Analysis 3) investigates the impact that different spatial configurations can have on continuous, equitable participation.

The overall research question, data sources and methods used in these three analyses are summarised in Figure 5.1, relative to the context of the whole thesis.

| | chapter 4 | chapters 5-7 | chapter 8 | chapter 9 |
|-------------------------|--|--|--|--|
| Question | <p>Q1</p> <p>How can the whole group engage with the topic?</p> <p>a) What types of engagement goals are desirable and feasible?</p> <p>b) What mechanisms can effectively support these goals?</p> | <p>Q2</p> <p>How should one orchestrate the activity?</p> <p>a) What constitutes the orchestration load?</p> <p>b) How should the orchestration load be balanced between teacher and technology to achieve feasibility?</p> | <p>Q3</p> <p>What are useful abstractions to aid future design, orchestration and evaluation?</p> <p>a) How can we generically describe the factors that make or break the design?</p> <p>b) What are practical recommendations for teachers and designers?</p> | <p>Q4</p> <p>Are these abstractions also useful when designing for sustainability?</p> <p>a) Is the CLOVER framework applicable in different settings?</p> <p>b) What design factors are critical for achieving sustainability in the wild?</p> |
| Study | 4Decades (lab) | 4Decades in the wild | | UniPad Comfy Birds |
| Data | lab video, field notes system logs teacher interviews student questionnaires debriefing transcripts | in-situ video, field notes system logs, screenshots teacher interviews student questionnaires debriefing transcripts | findings from 4Decades and literature | notes from design meetings with teachers teacher interviews field notes system logs |
| Analysis methods | lab-based prototyping qualitative analysis | detailed video analysis qualitative analysis of system screenshots | reflection on design development of a conceptual framework | lab based prototyping preliminary feasibility evaluations in the wild qualitative analysis |

Figure 5.1.: Chapters 5-7 in the context of other chapters.

5.1.1. Three analyses, four sessions, one study

To avoid confusion, it is worth disambiguating what I mean by study, analysis and session, respectively.

1. The entire work with 4Decades (Chapters 4-7) is referred to as one *study*, in contrast to UniPad study and the ComfyBirds study (Chapter 9).
2. As part of the 4Decades study, three different *analyses* were conducted, as explained above, corresponding to Chapters 5, 6 and 7.
3. The data came from four independent training course *sessions*. Each session took place on a different day with a different group of participants. The sessions lasted between one and two hours and were facilitated by the same teacher.

Figure 5.2 clarifies which parts of the data were used in each analysis. Analysis 1 (shown in yellow) draws solely on data from session 1, looking at the session as a whole, including the introduction, gameplay and debriefing. Analysis 2 (blue) focuses on the *rule-changing* feature that was first trialled in session 2. Finally, Analysis 3 (purple) compares data from all four sessions, while focusing on the *game play* part of each session.

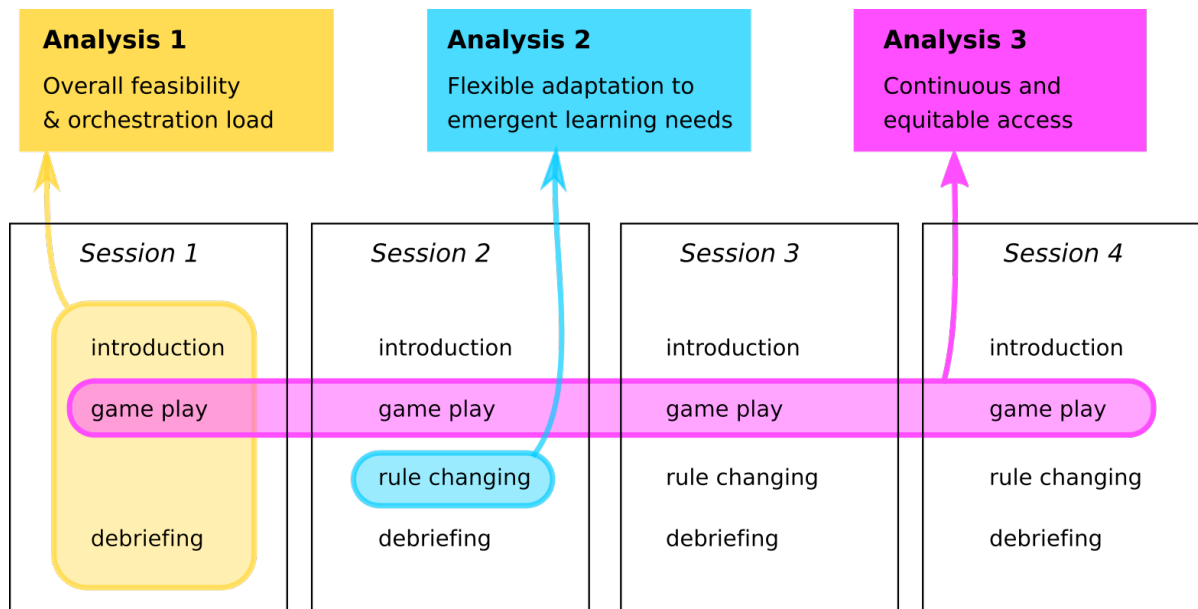


Figure 5.2.: Three analyses focused on data from four sessions.

5.1.2. Research aims and method

The overall research question for the entire 4Decades study can be summarised as:

■ What factors are instrumental in orchestrating such activities in the wild?

Each analysis addressed a different sub-question, as explained in the respective chapters. The overall aims were as follows:

1. to evaluate the 4Decades application in a real-life educational context;
2. to continuously test and improve the design until the application was mature enough to consistently satisfy the Gameplay Objectives;
3. to comprehensively describe the teacher's orchestration load;
4. to provide empirical validation for the instructional approach suggested by the literature review;
5. to provide tentative recommendations for generalisation and future design.

The analytical methods were mostly qualitative, drawing on video, participant questionnaires, system screenshots and partial transcripts of in-situ sessions. Some of the video and transcripts were also used in semi-structured interviews with the teacher. The qualitative analysis was partly supported by whole-session visualisations of relevant group interactions and the teacher's behaviour, as explained further below.

5.1.3. Setting

All four sessions took place in Cambridge, UK, in 2011. The locations were conference centres (sessions 1, 3 and 4) and a multi-purpose room on a college campus (session 2). In each session, a fixed time slot was available for the activity, including introduction and discussion. This amount of time was fixed (between one and two hours - different in each session) and without any leeway for going overtime. Setting up the technology in advance took roughly two hours per session.

5.1.4. Technical implementation

The application was implemented using a client-server architecture, similar to a dynamic website. The structure is outlined in Figure 5.3.

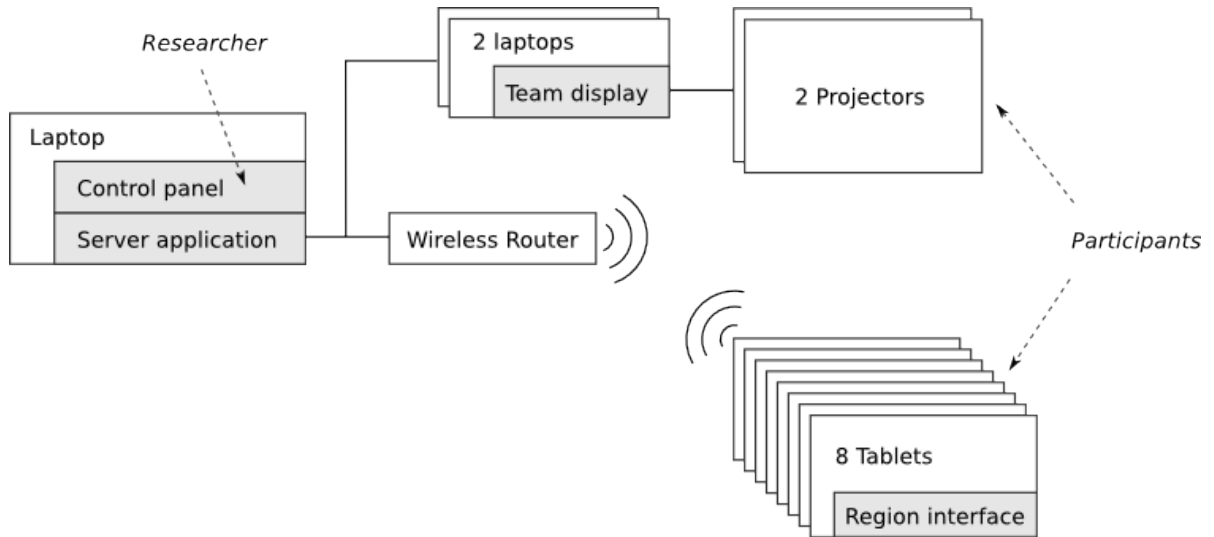


Figure 5.3.: Overview of hardware (white) and software (grey)

The application comprised the following software components:

- **Server application.** Implemented as a Java web application, this part contained the climate model and allowed the other devices to connect via a wireless router in the room. To ensure appropriate network reliability and response times, we brought our own router, rather than relying on the conference centre's local IT services.
- **Control panel.** The same laptop that hosted the server application also hosted a bespoke graphical user interface for monitoring and controlling the simulation. A researcher (myself) used this interface during the session. It served two purposes: (a) to verify good network connectivity of all devices and (b) to move the simulation forward to the next stage, in response to the teacher's verbal cues.
- **Team display** This Java application used data from the server application to visualise via a projector. Two instances were used, one for each team's ambient display, each on a dedicated laptop that was connected to the server.
- **Region interface** The browser-based interface for the tablets was implemented in HTML5, CSS and JavaScript. Using the default browser (Safari) in full-screen mode, the interface was usable like a native app. In contrast to a native app, however, this solution meant that no software needed to be installed on the devices - rather, the browser pulled the software automatically from the server. This was a minor practical advantage as most of the iPads were borrowed. More importantly, though, this solution guaranteed that the software was always up to date on all devices, which reduced the risk of errors.

To ensure equal chances for all participants, identical hardware models were used for the projectors and all the tablets (first generation iPads).

5.2. Analysis 1: Overall feasibility

The rest of this chapter reports on the first analysis, addressing the question:

Q2-A. Is it feasible to orchestrate the activity in real training courses?

The analysis was based entirely on session 1 (left part of Figure 5.4) which took place in a conference centre in the afternoon of the second day in a four-day CLP training course. The participants were 26 course attendees and some faculty.



Figure 5.4.: The following analysis focuses on session 1

In the following subsections I will describe the methodological approach taken in this analysis, outline the structure of the session in terms of the physical room layout and an overview of the teacher's actions, followed by a teacher-centric transcript of the introduction and practice round. After this overview, I will present the main findings, grouped in five categories: (1) the groups' decisions and outcomes in the simulation, (2) participant satisfaction, (3) evaluation of the Gameplay Objectives, (4) unexpected observations, and (5) a descriptive account of relevant tendencies in group interaction. A discussion of feasibility and orchestration load concludes this analysis.

The main conclusion from Analysis 1 is that the *interaction design* was indeed feasible in the wild but the *instructional design* offered potential for improvement. This potential was addressed in Analysis 2 (next chapter).

5.2.1. Method

To test the overall feasibility and investigate the overall orchestration load, the analysis considered the whole session from beginning to end; looking at group interaction on the whole-classroom level as well as teams, sub-teams and individuals. This implied multiple stages of video analysis, mainly serving two purposes:

1. to identify critical incidents (breakdowns and breakthroughs) and examine their causes, using detailed qualitative analysis of group interactions;
2. to create a high-level overview of the teacher's and groups' actions, by coding relevant behaviours at fixed intervals over the course of the entire session.

Besides video, other data sources were screenshots from the ambient displays, partial transcripts of group interaction, a full transcript of the debriefing and a brief post-hoc questionnaire asking participants to rate the overall activity on a 5-point Likert scale and allowing written comments.

Some of the data, particularly some parts of the video and screenshots, were discussed together with the teacher. This was done to gain better insight into (a) teacher-centric aspects of usability, including the orchestration load, and (b) the teacher's expert opinion on relevant learning needs and how effectively these were addressed by specific aspects of the activity. Direct suggestions for improving the design were also elicited.

To guide the feasibility evaluation, the 9 Gameplay Objectives from the previous chapter were used as a starting point.

5.2.2. Session structure

The session began with the teachers' introduction that took 15 minutes and included a practice match. The gameplay phase included 4 "proper" matches. From the moment participants entered the room until the end of the 14-minute debriefing the session lasted 67 minutes (Figure 5.5). Since each match comprised 4 rounds, 20 rounds were played in total, shown as black rectangles in the figure. The total duration of rounds was 33 minutes, i.e., the tablets were interactive roughly 50% of the whole session.

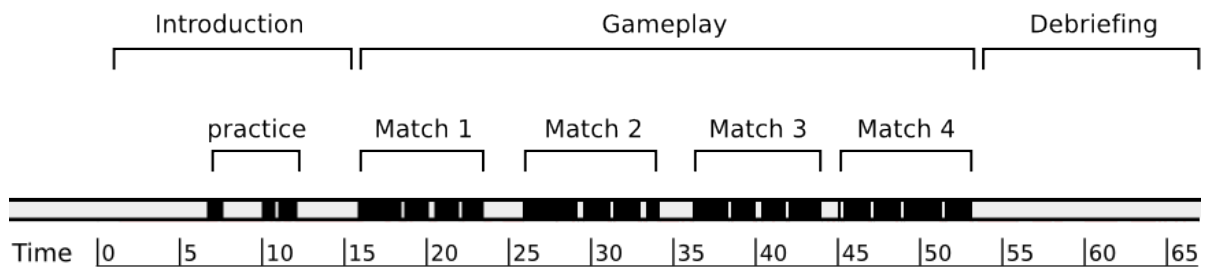


Figure 5.5.: Timeline overview of session 1

5.2.2.1. Room layout

Figure 5.6 shows the room layout in this session. The original layout is shown on the left hand side of the figure. This layout was changed, however, since the room was slightly narrower than expected and crowding was anticipated in the middle area near the ambient displays. To counter the risk of congestion, one display was moved over to the other side (right side of the figure.)

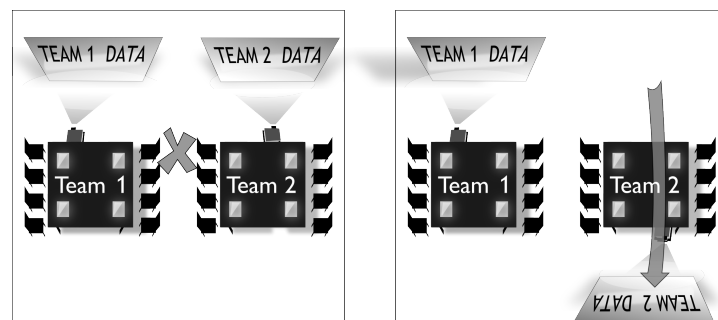


Figure 5.6.: Room layout in session 1

5.2.2.2. Overview of the teacher's actions

A summary of the teacher's actions during the session is shown in Figure 5.7.

Each of the 8 rows below the timeline (called “tracks” from now on) is dedicated to one of 8 actions performed by the teacher. The actions are themes identified through prior quantitative video analysis. Whenever the respective action was observed within a time window of 10 seconds, then it is indicated with a coloured rectangle at the respective point along the horizontal time axis. Conversely, white space in a track indicates that the behaviour did not occur in the respective time window. The occasional checkered (grey) rectangle represents missing data, e.g., brief moments when the teacher moved outside the view of the camera. The row below the tracks marks the beginnings of each

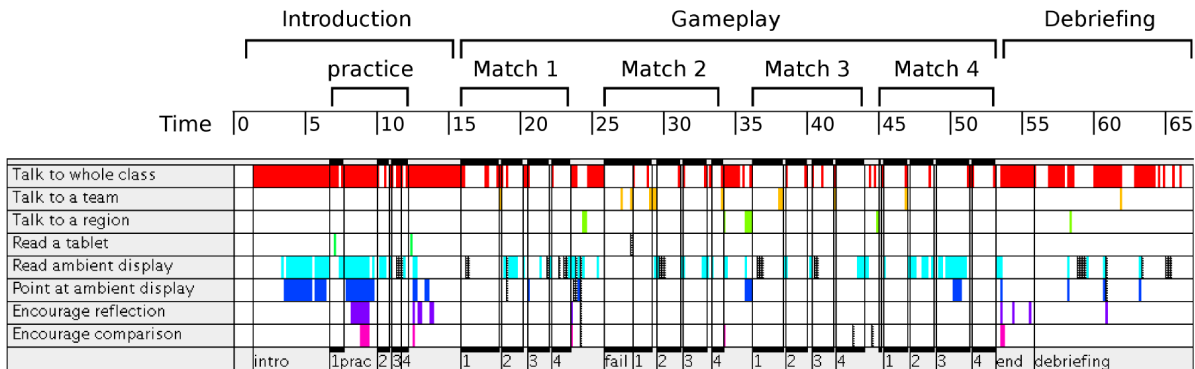


Figure 5.7.: Summary of the teacher's actions

new phase in the activity, i.e., the beginning of the introduction (“intro”), the four rounds of the practice match (“1 prac, 2, 3, 4”) that was part of the introduction, the four proper matches, the end of the last match (“end”) when the winner was declared, and finally the debriefing. The “fail” mark at minute 26 marks a technical interruption due to a temporary network problem with one of the iPads (which fortunately I was able to fix within less than two minutes). The start and end times of the rounds are marked with vertical lines and black rectangles above and below the tracks.

This type of visualisation will be used a few more times in this chapter and once in Chapter 9. Just to practice, let us consider the first track (red) that shows when the teacher was verbally active, addressing the class as a whole. At a glance one can distinguish the introduction, gameplay and debriefing in terms of the amount of teacher’s talk:

- **The introduction** appears as a continuous stream of talk, except for a few pauses during the practice rounds (as groups were then discussing amongst themselves).
- **During gameplay**, the teacher made only occasional announcements (mainly starting and closing the rounds) and talked more *between* the matches than during the matches.
- **The debriefing** was a back-and-forth dialogue between the teacher and the group.

Another thing the figure illustrates clearly is how the teacher used the ambient display(s). In the introduction the ambient displays showed a screenshot of the tablet interface, which the teacher “walked through” by explaining each element. These explanations were accompanied by frequent pointing to the display (see the blue areas starting at 3 minutes). After the first practice round (minute 7.5) the teacher walked through the teams’ decisions and results in a similar way, by pointing and explaining each number. During this procedure, which lasted two and a half minutes, the teacher encouraged

the group to reflect on the results (purple area). Walking over to the other display, he also encouraged them to compare one team's decisions and results with the other team's (pink). A similar use of the ambient display was observed almost consistently after each match and occasionally in the debriefing.

Whereas all the matches were roughly the same length, the rounds varied in duration, as the teacher paced the activity with a sensitivity for both the overall session schedule as well as the groups' varying need for decision time.

Finally, the figure shows that, by far, most of the teacher's actions were on the whole-class level, rather than individual teams or regions. The few rectangles in the "talk to a team" track represent very brief utterances, such as "have you finished making your decisions?" Similarly, the four occurrences of "talk to a region" during the whole gameplay phase represented rare moments when a participant asked a question. Remarkably, only twice did the teacher look briefly at a tablet screen (the first time to casually verify that the round had started and the second time to see which team won.) Both times lasted less than 5 seconds. Participants' use of the tablets never required the teacher's assistance, allowing him to focus on the group as a whole.

5.2.3. Introduction and practice match

A full transcript of the teacher's announcements during this time is represented below. The purpose is to give the reader (a) sufficient insight into participants' initial level of information and (b) a concise and exhaustive explanation of how the simulation works from a procedural point of view.

To ensure a fair distribution of roles, an equal number of red and green badges had been prepared and randomly assigned to the participants in advance. One group would initially be the players and the other group would be the analysts. The roles were swapped after each match.

Wearing their red and green badges, all participants simultaneously entered the room in which the simulation was set up (Figure 5.8).

0:46 Fantastic, now those of you with the green badges, would you like to sit at the tables - in pairs - each of you grabbing an iPad - and the rest of you with red badges in your team, stand around them. **1:23** Are the green people sitting at the table? Green people, you are the players for this round. And the red people, the red people with the dots, you are the analysts - for this round - you are observing. **1:38** And



Figure 5.8.: Participants entering the room (left and centre). Pairs taking seats at the tablets, some faculty and researchers watching in the back (right).

don't worry because the analysts and the players are going to swap around. You'll get equal turns. **1:44** This is a climate change game. And the way that you interface with this game is through the iPads. **2:06** Erm, now. The best way of describing what this game is and how to play it is to give you a little practice go. So you can kind of just have a see how it works. So that's what we're going to do very very shortly. **2:21** Erm, before though, just listen carefully to this: This game mirrors some of the learning that you've done in the course so far. It's a simulation of climate change over four decades - the game is called 4Decades - from 2010 to 2050. And you have to kind of make decisions about adaptation and mitigation as you go through that time period and you're going to play several times. **2:49** It's about making money. It's about the planet which arrives in 2050 with the most - the highest - gross planetary income. That's how you win. **2:58** But in a way, the way that economics works - and this is actually true - there's a kind of an assumption that there's growth in the system to begin with. So it isn't really about *making* loads of money. The way you win this game is by not losing it. The game is yours to lose. **3:18** Okay? **3:20** So let's now, erm, have a look at the interface in front of you. I'm going to just go over to this screen. **3:25** This is what you see on your iPad and I'm just going to take about three or four minutes, just to kind of go through this. Can you, you can also see it on that one over there. **3:34** At the top, you see, we begin the game in the year 2010. And you'll see the next line is GRI, gross regional income. Each of the iPads on your table represents a region on your planet. So each planet has four regions. And because you're looking at your region - which is your pair, or your iPad - it tells you what the income is for 2010 in your region. Now you all start - all regions on both planets start with 18 trillion dollars, which is calibrated to today's value of 70 billion PPP, that's where we start. **4:17** You also notice here - this is what your regional CO₂ emissions are right now in 2010. **4:23** Now here you've got a couple of buttons which you're gonna have a play with in a minute. One is the mitigation target for that

decade. And another one is called adaptation funding for that decade. **4:33** So this is your interface. This is where - you really only have a couple of things that you can do in this game. You can - you can add some mitigation or you can add or subtract some, er, adaptation. That's how you play the game. **4:46** You'll notice down here in this bottom part of the screen it says forecast. This isn't the kind of weather forecast. The forecast is what it actually will be. So whatever it says, at the end of the round it will be that. It's accurate. **4:58** And it says, you know, if you do this - or don't do this - this is what your CO₂ emissions will be a decade from now. So that would be in this case 2020. And according to what you've done this what the global temperature will be on your planet in 2020. **5:18** And there's a couple of lines here which relate to something we talked about this morning and yesterday. About climate damages. The first line, it says "baseline climate damages". If you do not invest in adaptation then you're gonna face - the planet warms, there's a certain amount of cost associated with that - we're uncertain but there is a cost. In this game we're not uncertain, it's got a distinct function. And if you don't do anything you're, er, regional economic growth at the end of this decade will be reduced by this amount. **5:59** If you *have* invested in some adaptation, your *residual* climate damage - that is the damage once we've taken into account how much you've invested - should be lower than that if you spent some money on adaptation. So that's where you get the difference between baseline damage and residual climate damage. **6:17** And at the bottom is your forecast. This is a key one. This is if you make those decision, this is what your gross regional income will be at the end of the decade. Okay? **6:27** So far so good. Okay, let's have a practice. **6:35** If you haven't got an iPad in front of you, just pair up, just have a look over the shoulder. Can we just turn the temperature down? Or is it just me having a hot flush? **6:48** Let's have a game, shall we. Now this is just a practice. This is purely for fun. Don't worry. **6:57** Now, here we are and your iPads should say 2010. And I'm gonna give you just a short time to make your decisions - how you're gonna invest in mitigation and adaptation for the next decade off you go.

[round 1 starts]

7:30 (What can you see, what do you notice) **7:35** Okay, as this is a practice, I'm gonna call that round to a close, now

[round 1 closes]

7:49 Right, if I could have your attention? If I could have your attention. If you see over here - Betelgeuse - we just had a round. A round is a decade. Okay? A round

is a decade. And in Betelgeuse - just see here - these are the four iPad selections on Betelgeuse. 2, 3, 2, 1 is what the regions spent on mitigation. The region that selected 2 gigatons as a goal at the end of the decade selected 1 trillion dollars of investment in that region. So you can see that's paired up. That's the individual iPad choices for mitigation for and adaptation. **8:33** And consequently the emissions in those regions are slightly different because you didn't all choose the same mitigation target. **8:39** You have one temperature on your planet. Because it's one planet. Yeah? And it's global meaning - that's not true, I don't want to tell you that but - we talk about global mean surface temperature, this is what this is. So on Betelgeuse it's 1.2 degrees. What temperature is it over on Alpha Centauri? 1.1. The baseline damage is this, okay? And that's the same because that - you would all suffer that if you hadn't invested. **9:04** But some people invested in adaptation. These regions, you see, don't have any residual damage because they've done some investment. **9:12** Region 3 - if I can call it region 3 - did no adaptation and were suffering a little bit of residual damage. And in the gross regional income each of the four iPads in this planet is here. Yeah? So at the moment region - this first one - you don't actually know who you are - but you will if you compare the score on your thing. **9:35** So what you'll see - analysts, you're going to play this game, too. And players, you're going to be analysts. When you're not playing the game - when you're not paired up - your job is to pay attention at the end of each round [points to ambient display] **9:54** what are you learning from these screens? How do we play this game? What's the best strategy? How do we maximise our planetary income. Let's carry on very quickly, and we'll go through, er, three more decades very quickly because this is a practice, okay, off you go.

[round 2 begins]

10:13 Take your choices for the next decade. [photo taken, see Figure 5.9]



Figure 5.9.: The first cohort playing a practice round

10:32 Okay, I'm gonna rush you along. This is just a practice. Erm, another few seconds. **10:55** Are you done? Okay - round two is closed!

[round 2 closes]

10:59 Now we're in 2030. Look at what you see. What do you see? Look at your - Let's play round three.

[round 3 begins]

11:07 Let's go to 2040. Now you're going to make some more investment decisions.

11:30 And let's bring round 3 to a close.

[round 3 closes]

11:40 Let's play round four. Let's go up to 2050 in this practice round.

[round 4 begins]

12:06 And let's bring the game - let's bring the match, the match to a close. Are we closing the match?

[round 4 closes - end of match]

12:15 And what do you see on your iPads? What does it say? What does it say on your iPad? [peeks onto one of the tablets]. Who won? **12:25** Betelgeuse won. Look, Betelgeuse achieved a GPI - gross planetary income - of 110.1 trillion. Versus Alpha Centauri's 54.5. **12:45** Okay. So, did you notice - can I just have your attention - did you notice, as you were choosing - as you were sort of looking at your iPad and thinking about what to choose - that forecast was kind of changing in real time wasn't it? Even when you weren't playing with it. That's because you're a planet. You're a team and *you're* a team. And even if you as a region are saying "I don't know what to do", your neighbours they are playing with the buttons and they're changing the outlook for you. That's why that changes in real time. **13:24** Okay, now I want to make it very clear how you win this game. What we have just played is a match. And a match is a simulation from 2010 to 2050. And I took you through it very quickly. I'll give you a little bit more time for the next matches as we go along. Erm, you may have noticed - or you may have heard this morning or yesterday - that when you make investments in adaptation they don't come free. There is that marginal abatement cost curve. And there is that stuff I showed you at the end of my presentation, which is about the net benefits of adaptation. It will cost you something. So in this game - this game kind of mirrors that reality. And on the table you'll see a piece of paper which is your - with 4Decades written on it - and that is your adaptation and mitigation cost

curve. I realise there's only one piece so you may pass it around and have a look. And I'll give you a minute or so to study those. Now, part of the game is understanding - oh no, is there several sheets? Oh yeah, pass them round, pass them round. Okay. **14:30** When you make your decisions, bear in mind that mitigation cost curve. And bear in mind the cost-benefit of making adaptation decisions. Notice on the adaptation cost-benefit structure, notice that that changes as a function of temperature. It isn't always the same. **15:23** Okay, now, the winner of today's game will be the team - Alpha Centauri or Betelgeuse - that achieves the highest score over four matches. [...] What do you win? You're gonna win the champagne over here. Which will get you started early for the next session. **15:30** Is that clear? Are there any questions? Shall we begin!

5.2.4. Findings

The following subsections report on the findings, including the teams' decisions in the simulation, overall participant satisfaction, verification of the Gameplay Objectives and unexpected observations.

5.2.4.1. Team decisions and results in the simulation

Participants showed a good understanding of how to play the game. By hypothesising, discussion, using ambient information and coordinating their decisions, both teams figured out strategies that led to a higher GPI (i.e., gross planetary income, which was defined as the winning criterion). Both teams' improvement from the first match to the last can be seen in Figure 5.10.

5.2.4.2. Overall participant satisfaction

Participants were engrossed in the activity from the beginning of the game and stayed engaged throughout the 90 minutes session. In a paper-based post evaluation, the whole session was rated on a Likert scale, from 1 to 5, with an average score of 4.1, which was close to the whole course's average rating of 4.23. Out of 20 responses no-one gave the lowest score of 1. A further indicator of success was receiving several invitations from delegates to deploy 4Decades in their organisations. Moreover, the course committee expressed strong interest to use the simulation again in future courses.

Alpha Centauri

| MATCH 1 | | Final GPI 176.2 | | | |
|---------|---------|-----------------|------|------|------|
| Mitig | Adapt | GRI | | | |
| 2 2 2 2 | 1 1 1 1 | 23.2 | 23.2 | 23.2 | 23.2 |
| 1 1 1 2 | 2 2 2 2 | 31.1 | 31.1 | 31.1 | 30.1 |
| 2 3 1 3 | 3 2 2 3 | 40.0 | 39.0 | 42.0 | 36.6 |
| 3 2 0 3 | 4 2 4 4 | 44.1 | 43.1 | 48.2 | 40.6 |
| MATCH 2 | | Final GPI 174.9 | | | |
| Mitig | Adapt | GRI | | | |
| 4 1 3 3 | 2 1 2 2 | 16.2 | 24.2 | 20.2 | 20.2 |
| 2 2 3 3 | 4 2 2 1 | 18.4 | 31.4 | 23.9 | 24.7 |
| 1 2 1 1 | 1 2 1 1 | 25.5 | 41.5 | 33.1 | 34.1 |
| 1 4 1 1 | 2 2 2 2 | 34.2 | 49.5 | 44.8 | 46.2 |
| MATCH 3 | | Final GPI 214.3 | | | |
| Mitig | Adapt | GRI | | | |
| 1 1 1 1 | 1 1 1 1 | 24.2 | 24.2 | 24.2 | 24.2 |
| 1 1 1 1 | 1 2 1 1 | 33.4 | 32.4 | 33.4 | 33.4 |
| 1 1 1 1 | 2 2 2 2 | 45.3 | 43.9 | 45.3 | 45.3 |
| 1 1 1 1 | 9 9 9 9 | 53.9 | 52.5 | 53.9 | 53.9 |
| MATCH 4 | | Final GPI 215.7 | | | |
| Mitig | Adapt | GRI | | | |
| 1 1 1 1 | 1 1 1 1 | 24.2 | 24.2 | 24.2 | 24.2 |
| 1 1 1 1 | 1 1 1 1 | 33.4 | 33.4 | 33.4 | 33.4 |
| 1 1 1 1 | 2 2 2 2 | 45.3 | 45.3 | 45.3 | 45.3 |
| 1 1 1 1 | 9 9 9 9 | 53.9 | 53.9 | 53.9 | 53.9 |

Betelgeuse

| MATCH 1 | | Final GPI 198.4 | | | |
|---------|---------|-----------------|------|------|------|
| Mitig | Adapt | GRI | | | |
| 1 2 2 2 | 1 1 1 1 | 24.2 | 23.2 | 23.2 | 23.2 |
| 1 1 1 1 | 1 1 1 1 | 33.4 | 32.1 | 32.1 | 32.1 |
| 1 2 3 1 | 2 2 2 2 | 45.3 | 42.4 | 40.4 | 43.4 |
| 0 1 3 3 | 8 5 6 7 | 54.3 | 49.3 | 45.3 | 49.4 |
| MATCH 2 | | Final GPI 169.9 | | | |
| Mitig | Adapt | GRI | | | |
| 4 3 3 4 | 1 0 1 1 | 17.2 | 20.6 | 21.2 | 17.2 |
| 2 3 2 3 | 1 1 1 1 | 22.8 | 25.5 | 28.2 | 20.8 |
| 1 1 1 1 | 1 1 3 3 | 31.5 | 35.3 | 37.0 | 26.7 |
| 1 1 1 0 | 2 7 3 2 | 42.6 | 42.9 | 49.2 | 35.0 |
| MATCH 3 | | Final GPI 194.1 | | | |
| Mitig | Adapt | GRI | | | |
| 2 2 2 2 | 1 1 1 1 | 23.2 | 23.2 | 23.2 | 23.2 |
| 2 1 2 2 | 1 1 1 1 | 31.1 | 32.1 | 31.1 | 31.1 |
| 1 1 1 1 | 3 3 3 3 | 41.0 | 42.4 | 41.0 | 41.0 |
| 0 0 0 0 | 8 7 3 7 | 48.7 | 50.7 | 45.3 | 49.3 |
| MATCH 4 | | Final GPI 204.8 | | | |
| Mitig | Adapt | GRI | | | |
| 3 3 3 3 | 1 1 1 1 | 21.2 | 21.2 | 21.2 | 21.2 |
| 2 2 1 3 | 1 1 2 1 | 27.9 | 27.9 | 28.3 | 25.9 |
| 1 1 1 1 | 1 1 2 1 | 38.4 | 38.4 | 38.2 | 35.9 |
| 1 1 1 1 | 2 2 2 2 | 52.1 | 52.1 | 51.8 | 48.6 |

Figure 5.10.: Excerpt of the ambient displays, showing both teams' decisions (mitigation and adaptation) and results (GPI = gross planetary income). Some columns are omitted to save space (zig zag lines).

5.2.4.3. Findings regarding the Gameplay Objectives

The session was a success in the sense that each of the 9 Gameplay Objectives was satisfied to a reasonable degree. Findings relevant to each objective are presented below.

Objective 1: Allow participants to apply relevant course knowledge in practice through decision-making and debate

To get better results in the simulation, groups had to operate strategically with relevant concepts from prior lectures, such as mitigation, adaptation, emissions, damage, etc., and to negotiate decisions as a teams. This provoked continuous and meaningful discussion on the topic.

The groups were engrossed in these discussions, focusing on the activity from beginning to end. The only instance where a participant disengaged from the activity was during one of the matches, as one participant stepped aside to use their phone for a minute, before rejoining the group. Apart from this brief incident, everyone seemed constantly immersed in the social situation. All conversations at both tables were consistently on topic, including the joking. This observation was based on high-definition microphones on both tables, that were capable of picking up multiple lines of simultaneous conversation at reasonably intelligible quality. Although the microphones failed to cover conversations that took place further away from the tables, participants in those areas were also found to be focused on the simulation content, as could be inferred from their body language, such as frequent pointing and looking at the ambient displays and tablets.

Objective 2: Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use

As the teacher initiated the first round, all eyes were on the tablet devices. Seated pairs were immediately engrossed in pressing the buttons and watching the numbers change, thinking aloud together while exploring the interface.

The “analysts” (i.e., the standing participants) watched the tablets over their teammates’ shoulders. It appeared that the tablet interfaces were immediately accessible and understood by participants. Sharing the tablets worked seamlessly as tablets were kept centred between pairs and participants naturally drew their hands back after pressing

buttons, thus effectively releasing control and freeing the screen for others to peek in (Figure 5.11).

The analysts seemed to benefit from this affordance as they often used their mobility to peek into several tablets to assist collaboration from more of a bird's eye view. Standing participants were also engaged in dialogue with pairs. They contributed mostly in verbal form, e.g. suggesting theories and strategies, directing attention to other regions or data on the ambient displays, but rarely touching the tablets directly.



Figure 5.11.: Touch-screen devices were found easy to share between two players. Analysts were able to peek in over the shoulder.

No usability issues regarding the tablets or ambient displays were observed or mentioned in the debriefing or questionnaires. All participants had a clear view on their own team's ambient display and some could read both ambient displays from where they were sitting or standing.

Objective 3: Allow a single teacher to facilitate and debrief the activity, using an ecology of displays

Besides the obligatory introduction, debriefing, etc., one of the teacher's responsibilities was to manage the pace of the session by opening and closing the rounds. Between the rounds it was easy for the teacher to get the whole group's attention, since the iPads were then in a non-interactive mode. Sometimes the teacher used these moments to give hints, interject bits of theory, make comparisons with real life global policy, clarify a player's question or encourage helpful practices such as the use of pen and paper.

The ambient displays were useful in helping the teacher participants stay abreast with the teams' decisions. Initially the teacher compared between the teams' decisions and outcomes, thus encouraging awareness between the teams.

During the rounds the teams were engrossed in game-play without needing the teacher's assistance. Questions were typically sorted out among peers with the effect that the teacher never had to divide his attention at any point in the session. This allowed the teacher to stay in the background most of the time, casually observing the whole-room interaction and the data on the ambient displays.

At the end of each round, when the tablets switched to a non-interactive state and the data was updated on both teams' ambient displays, the overall attention of the teams shifted noticeably from the tables to the whole room. The teacher occasionally used these breaks between the rounds as opportunities to get the attention of the whole class, e.g. to provide additional information, without interrupting the flow of the game.

The only time that the teacher looked at one of the tablets was to verify who won after the practice match. Later he deduced this information by comparing the final result on the ambient displays. This was less necessary towards the end as teams had become nimble at reading the results and their exultant cheering gave the winner away instantly.

Objective 4: Encourage participants to reflect on assumptions through decision making and discussion

Based on prior lectures, some participants had apparently assumed that high initial spending on mitigation would pay off best in terms of final income:

- *“It's a climate change game. The answer has to be: mitigate - a lot - early.”* (Team B, 2nd match, 28:38)

Several participants commented in the debriefing that they were surprised how small the payoffs for high initial mitigation spending were.

- *“We were slightly surprised that a bit more spending on mitigation in the early decades didn't deliver a better result, which is what you'd have thought from what we were hearing and discussing today. So that was a bit counterintuitive.”*

Some suggestions indicated that the simulation would better reflect their prior expectations if the economic impact of global warming was higher.

- *“I didn’t think there was enough penalty. We were trying to think well actually, trying to manage temperature earlier, but that wasn’t coming through in terms of the results.”*
- *“Yeah, the other thing that’s surprising is that the temperature rise at the end doesn’t - you’d expect it would start hurting you much more, in damage, than it actually does. The temperature rise was, you know 5.2. That would damage your planet.”*
- *“You [the other team] should have been punished.”*

Objective 5: Allow participants to refresh from exhausting lectures by moving and talking

Every participant took part in the discussions. The group was moderately active in terms of talking and gestures. No signs of fatigue or physical discomfort were observed or reported. Switching roles between sitting and standing (approximately every 10 minutes) created opportunities for participants to move their legs a bit. However, the narrow room did not allow much strolling around and some of the analysts ended up staying in the same place, some leaning on the walls.

In contrast to the one-directional orientation in a typical lecture, participants frequently turned in different directions to read information, point, talk and listen. Although participants were not asked explicitly whether they found the activity refreshing, general enjoyment was indicated by the high frequency of laughter, joking and cheering, as well as some comments in the paper-based feedback forms: *“Great fun - course needed active participative session.”*, *“Good to include a fun learning event.”*, *“Really good break in the day.”*

Objective 6: Encourage negotiation by letting participants follow different goals but requiring the consent of teammates

The official winning condition had been unambiguously defined by the teacher as “highest income”. This was commonly understood by participants, as the dialogues and strategies showed, with both team evidently being keen to win. However, in both teams a minority of participants challenged the idea that money was the most desirable goal, and thus suggested minimising temperature instead. This fundamental disagreement provoked lively discussions which carried over into the debriefing. Two participants from team B later reflected:

- *“We made slightly less money than them, but considerably, erm we have, you know, saved the planet.”*
- *“The money was a driver, but, i think we didn’t, erm, we certainly could have made a lot of our decisions based on trying to keep the temperature down.”*

Team A consistently followed a money-driven strategy, although the suggestion to minimising temperature kept popping up.

- *“Look, the best temperature was 3.5 after match 2.”*
- *“Yes, if you got a bonus for that, it would be a different game.”*
- *“Don’t you care about temperature? Cynics!”*

As soon as teams had realised the implications of the mitigation cost curve, they tried to enter identical decisions on all the tablets. This was more successful in team A than in team B, as some participants noted in the debriefing:

- Team A: *“There was a lot of really good collaborative policy making going on, probably on both teams. Especially after the first [match] everybody started to kind of communicate across the table. And that fairly rapidly got us towards [a joint winning strategy]”.*
- Team B: *“I think what was interesting as well is, erm, the disagreement over kind of - you think you’ve made an agreement and then, you know, and then a region changes their opinion.”*

The suggestion that team A was more coordinated than team B is supported by comparing the decisions on the two ambient displays (Figure 5.12).

Objective 7: Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play

Collaboration initially happened within pairs, as players used the tablets to make decisions. The analysts, initially watching quietly, soon began to make suggestions and engage in hypothesising and discussions. As the roles within a team swapped after each match, small groups of players and analysts often stayed together, trying out local strategies before sharing them in the whole team. To help coordinate team decisions, some analysts temporarily took on mediating roles between two or more iPads, by standing or walking between them.

| Alpha Cen... | Betelgeuse |
|--|--|
| MATCH 1 Mitig Adapt 2 2 2 2 1 1 1 1 1 1 1 2 2 2 2 2 2 3 1 3 3 2 2 3 3 2 0 3 4 2 4 4 | MATCH 1 Mitig Adapt 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 1 2 2 2 2 0 1 3 3 8 5 6 7 |
| MATCH 2 Mitig Adapt 4 1 3 3 2 1 2 2 2 2 3 3 4 2 2 1 1 2 1 1 1 2 1 1 1 4 1 1 2 2 2 2 | MATCH 2 Mitig Adapt 4 3 3 4 1 0 1 1 2 3 2 3 1 1 1 1 1 1 1 1 1 1 3 3 1 1 1 0 2 7 3 2 |
| MATCH 3 Mitig Adapt 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1 1 1 1 9 9 9 9 | MATCH 3 Mitig Adapt 2 2 2 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 3 3 3 3 0 0 0 0 8 7 3 7 |
| MATCH 4 Mitig Adapt 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 1 9 9 9 9 | MATCH 4 Mitig Adapt 3 3 3 3 1 1 1 1 2 2 1 3 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 2 |

Figure 5.12.: Team A (left) was perfectly coordinated in the last match (green highlights at the bottom left), whereas team B (right) always had some disagreement (red).

Evidently, the teams were keen to outperform the other team, as evidenced through engaged discussion about winning strategies, cheering and clapping at the end of rounds and matches.

Objective 8: Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data

After each round many participants turned their heads to see their team's results appearing on the screen and compare them with the other team. This type of comparison was the main use of the ambient displays in the early matches. As the amount of data grew over several matches, participants became quicker at knowing what to look for, helping them reflect more readily on the aggregate data and theorise about the implications. Increasing numbers of participants spent time focused on their own team's or the other team's scores, discussing ideas in small sub-groups or pairs.

Some of these discussions happened as independent threads outside the main action at the tables. For instance, Figure 5.13 shows two participants studying the opponent's ambient display for an extended period in order to learn from the other opponent's strategy. Later they went back to their team and persuaded everyone to adopt the plan they had devised.

Furthermore, the ambient displays were used in the debriefing discussion to recall significant moments and support arguments with concrete data.

Objective 9: Allow all participants to contribute equitably to the activity and ensure that no-one is excluded

The role-switching guaranteed that participants spent equal time as players and analysts. Players were paired up at the tablets which were shared equitably. Despite the narrow space, analysts always had enough room to easily look into one or two tablets over the players' shoulders. Moreover, all players could easily see their own team's ambient display; and the analysts could usually see both displays without much effort.

5.2.4.4. Unexpected Observations

The above findings were related to the 9 Gameplay Objectives. In addition, the following four unexpected observations emerged, as detailed below.



Figure 5.13.: Two men in the foreground studying the other team's ambient information for more than a minute

1. Collaboration expanded and accelerated over time
2. A minor breakdown occurred with regard to team awareness
3. A new learning need emerged: Playing by different rules

Collaboration expanded and accelerated over time

Over the course of the four matches, the extent of collaboration could be seen to gradually increase from controlling the simulation via tablets in pairs, to regional tactics, to team strategies and whole-classroom exchange of information. In the first matches, most of the talking occurred between pairs sitting by the tablets, changing the values on the screen, while the others were either watching intently or talking to their closest neighbours. Decisions in the game were made on a local basis without consulting the whole team. Later matches were characterised by efforts to establish a shared awareness of the whole team's situation and a team-wide strategy, as groups realised that local decisions affected the state of the simulation for the whole team. Increasingly, participants made eye contact and verbal contact with team members across the table and standing participants

started walking between regions, comparing values between devices, making suggestions and engaging in clarifications and arguments.

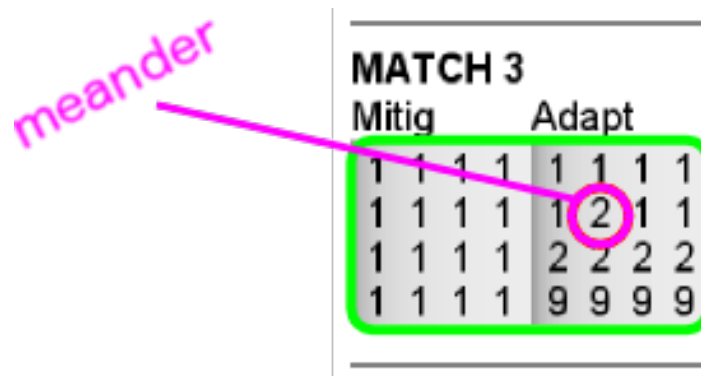
The groups continuously relied on the local, short-term control that the tablets provided. In addition, the global, long-term information on the ambient displays became increasingly important throughout the session. Ambient displays were found appropriate for persisting the large amount of global, long-term information about the state of the simulation and previous states, effectively freeing the tablet screens to contain only a minimum of information. The latter was found to be an important use quality, allowing participants to engage face-to-face, rather than having to constantly look down at the tablet screen.

Initially, teams' communication about decisions was rather verbose. Over the course of several matches, however, both teams independently developed abbreviations to coordinate decisions more quickly and efficiently:

- “1-9 everybody, 1-9!” (Team A)
- “Shall we all do 3-1 now! Are we together? Are we unified?” (Team B)

Limits of team awareness: The “meander” incident

Even when a team had agreed on a shared decision, this did not necessarily imply that everyone actually entered the same values. The following vignette describes a critical incident where one region “meandered” off to do their own little experiment. Although this happened only once (namely in Team A, Region 2, third match, see Figure 5.14), the story is worth telling, as it highlights some of the strengths and limitations of the chosen design for managing awareness among large groups.



| MATCH 3 | | | | | | | |
|---------|---|---|---|-------|---|---|---|
| Mitig | | | | Adapt | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 1 | 1 | 1 | 1 | 9 | 9 | 9 | 9 |

Figure 5.14.: The purple “meander” mark highlights an exceptional glitch in Team A’s internal coordination.

Figure 5.15 shows a storyboard-like representation of the vignette, whereby the yellow dots represent participants sitting at the table, blue are the analysts and green is the teacher. The black triangles indicate the dominant direction(s) of eye gaze. The team's ambient display would be depicted at the bottom of the picture, near the "main characters" A, B and C. In the following I will describe each of the frames. They are numbered 1-12.

1. After the team had quickly arrived at an agreement and entered identical decisions, the team is now waiting for the teacher to close the round. Meanwhile, some of the analysts are studying the ambient displays (in fact F and G are also depicted in Figure 5.13). Several conversations are taking place simultaneously, some talking amongst their region and some (including A) talking across the table. B is studying the team's own ambient display.
2. At some point B makes eye contact with A and asks a question regarding temperature, pointing to the tablet. C is looking at the tablet, apparently listening.
3. A follows B's gaze to the ambient display and looks back to the tablet.
4. A makes an observation regarding the temperature.
5. Pointing to the tablet, A suggests a strategic change for the next round (or the next match).
6. C makes a comment (that was not picked up clearly by the microphone). A and B make eye contact with C. A responds.
7. The teacher approaches the team with the usual phrase, offering to close the round shortly. Several participants answer yes and the teacher turns away to ask the other team. Meanwhile A, B and C have been immersed in their private experiment, apparently unaware that the round is about to close.
8. E notices that A, B and C are still experimenting with different numbers but does not intervene.
9. D realises as well and notifies A across the table that the time for decisions is up.
10. The round closes, rendering all the decisions permanent. A, B and C seem surprised by the unexpected end of the round. D is aware of the mistake.
11. A and B look at each other, seeming slightly puzzled. C seems to realise that their preliminary choice got locked in by mistake and differs from the numbers that the team had agreed on. The difference is apparent on the team's ambient display. More teammates realise the mistake, either through the ambient display or body language. Some are still busy studying the

other team's display (e.g., F and G). 12. Much laughter occurs and heads turn towards the "guilty" region, as the majority of the team seems to realise what is going on.

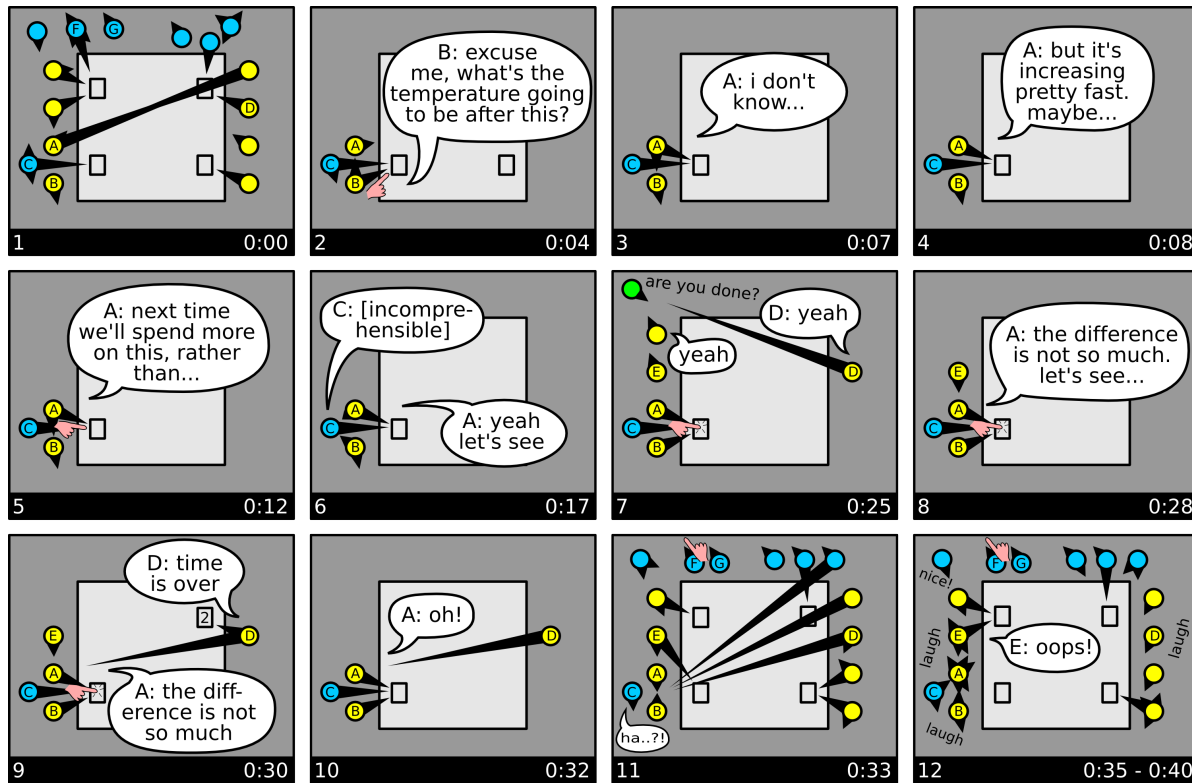


Figure 5.15.: While the team was waiting for the round to close, one region meandered off to do their own experiment.

A new learning need emerged: Playing by different rules

Several participants suggested, directly or indirectly, that it would be desirable to play the simulation by different rules - or making the rules variable. The following comments were made in the debriefing.

- "There's an anomaly because we have a cutoff of the decades."
- "Yeah, the other thing that's surprising is that the temperature rise at the end doesn't - you'd expect it would start hurting you much more, in damage, than it actually does. The temperature rise was, you know 5.2. That would damage your planet."
- "Within the game there's no price for getting it wrong on an aspect of it [...] 5.2 whatever it is, that would be a fairly toxic place to live I'm guessing."
- "You [the other team] should have been punished."

- *“Maybe we should factor in a maximum amount that you can adapt. Feasibility, technological feasibility.”*
- *“If there had been a fifth decade, I wonder what - how your team would have then fared.”*

Below are some comments made in the paper-based feedback forms.

- *“Very entertaining. I would do other rounds taking also temperature as a driver (not money).”*
- *“Great idea to embed learning via game but need to modify parameters so the winning strategy gives the right message on how to combat climate change.”*
- *“Needs more work. There should be consequences for higher temps.”*
- *“Parameters need to be tweaked from one match to another to avoid ‘copying’ of previous numbers used.”*

Overall, the critical comments concerned the game’s winning condition as well as the model parameters.

Tendencies in groups’ talk, gaze, and use of technology

To get an objective overview of interaction within teams, two regions were analysed in detail for talk, eye gaze and tablet interaction. The resulting visualisations are shown in Figure 5.16. For reference, the familiar diagram of teacher’s actions is also copied into this figure.

The following observations are supported by this visualisation.

1. Decisions were initially made within regions and later as a whole team. This is evidenced by the fact that “talk within region” (track 1) was high from the beginning, whereas talking across the table (track 2, “talk other region”) virtually didn’t occur at all during the entire practice match. By contrast, the “proper” matches 1-4 showed much talk across the table. This finding is consistent with a debriefing comment saying that *“especially after the first [match] everybody started to kind of communicate across the table.”* Arguably, this behaviour was encouraged implicitly by the teacher’s hint at minute 12:45 about the forecasts being connected between tablets (see above transcript of the introduction).

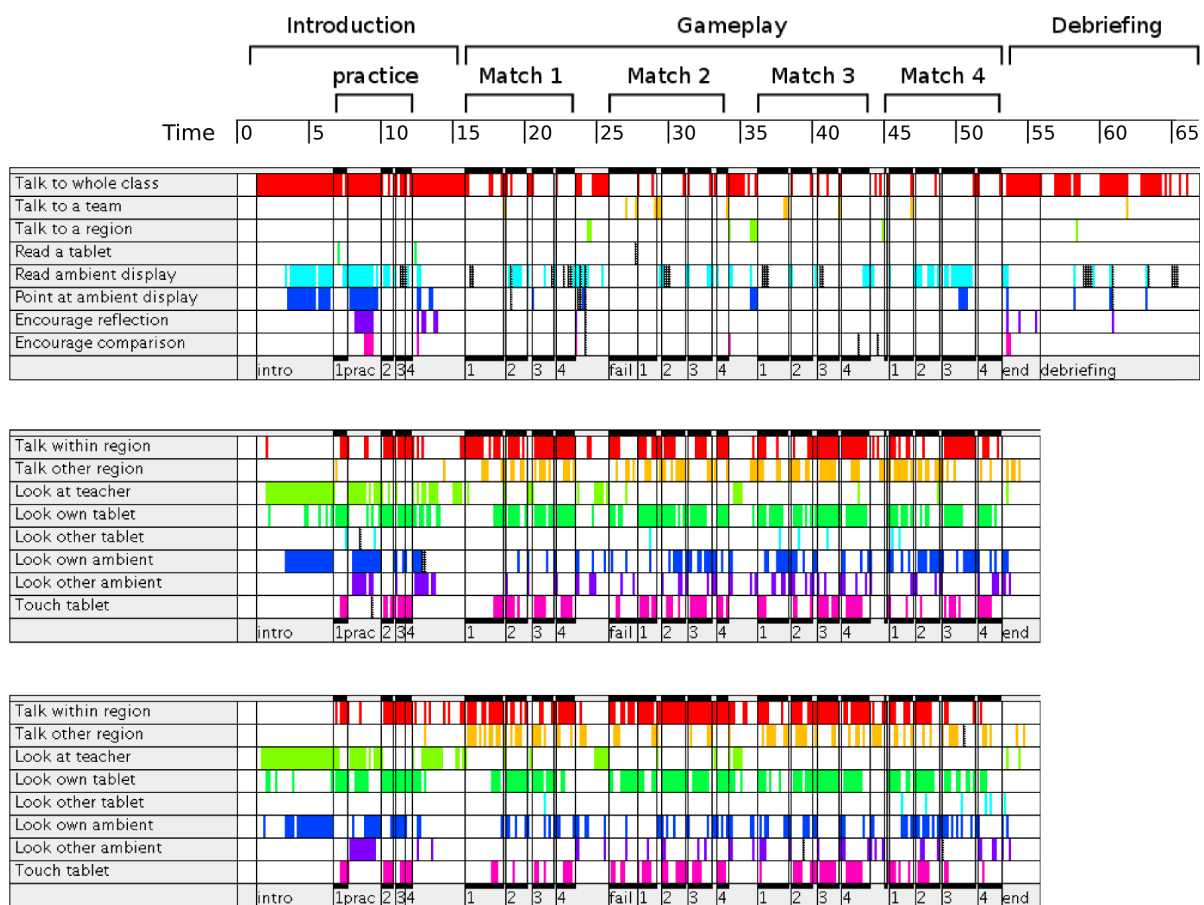


Figure 5.16.: Graphical overview of session 1: teacher (top row) and two regions in Team A (centre and bottom row)

2. Looking at the teacher (track 3) decreased remarkably during the gameplay phase. Whereas the teacher had participants' visual attention mostly during the introduction and in the first two breaks between the matches, one region was not observed looking at the teacher for the last 15 minutes of gameplay.
3. During the breaks between the matches, participants looked up from their tablets (track 4, "look own tablet").
4. Looking at another region's tablet (track 5) occurred sporadically, as analysts occasionally took on coordinating roles.
5. The ambient displays were consistently used at the end of every match.
6. Use of the ambient displays slightly increased over time.

5.2.5. Discussion

The research question for this analysis was:

Q2-A. Is it feasible to orchestrate the activity in real training courses?

5.2.5.1. Feasibility

The first in-situ evaluation of 4Decades went down without any major problems, suggesting that the activity was indeed feasible in the wild. All 9 Gameplay Objectives were met to a satisfactory extent. Participants generally described the activity as engaging, useful and enjoyable. High ratings and positive comments indicated that the course attendees truly welcomed the role-playing and felt that it added value to the lecture-based curriculum. This impression was shared by the organising committee who asked for follow-up deployments in future courses.

As far as the findings suggested at this point, the activity was considered stable and could feasibly be deployed in the wild again, without any design changes. However, potential for improving the design was suggested by the eventually unresolved dispute about how high the output values (temperature and damage) from the simulation should realistically be, given certain inputs. Since neither gameplay nor the debriefing provided optimal conditions for these discussions, an additional mode of interaction was devised, aiming to let participants change the rules of the simulation (Chapter 6). The rationale for this design iteration is detailed below. It depends strongly on the design's ability to

surface participants' tacit assumptions, enable subversive play, and encourage face-to-face conflict resolution.

5.2.5.2. Factors for inclusive game design

The following game-related properties were identified as conducive to achieving the Gameplay Objectives, particularly objectives number 2, 6, 7 and 9, regarding easy learnability and usability, familiar social mechanisms and inclusivity.

1. **Team play.** Team play combines the motivational benefits of collaboration and competition. The challenge of beating the other team (rather than beating the game) benefitted participants' *continuous awareness* of peers' actions and intentions. In regard to the specific topic, this was appropriate for reinforcing the notion that climate change is about people, rather than mere numbers.
2. **Non-oppositional game mechanisms.** In the simulation, the teams had no way of influencing each other, such as by trading, blocking or sabotage. This kept the game rules simple, while leaving room for face-to-face interaction (e.g. talking and spying on the displays). The latter were found sufficient for sustaining mutual awareness.
3. **Decision replacement over decision-making.** Since the input values defaulted to valid actions, (i.e. zero spending), the simulation assumed this as the group's decision unless they changed it explicitly. I.e., the interface did not enable abstention. This subtle feature had two important implications. Firstly, the default values provided a concrete starting point to trigger initial discussion. Secondly, all teams were guaranteed to have some usable output at all times - rather than some tasks being unfinished. This helped the teacher move the activity forward at agreeable moments, thus staying in control of the overall pace.
4. **Incremental changes.** The plus/minus buttons afforded incremental changes, rather than absolute values. This arguably lowered the entrance barrier for groups, by allowing them to approximate a desired result through small (and easily reversible) changes. Taking responsibility for such incremental changes is arguably less daunting than entering absolute values. Moreover, plus/minus buttons (as opposed to, e.g., sliders) are beneficial insofar as they do not per se suggest any "normal" values or limits which could potentially bias participants' judgement.

5. **An easy first choice.** The default values were easily recognisable as poor decisions. This may have benefitted participants' confidence to make changes, since they could conclude from prior lectures that a little spending was probably better than no spending.
6. **The impossibility of NOT communicating.** A related implication was that groups were thrown into a situation where they could not "not communicate" (Watzlawick et al. 1967). This meant that participants did not have to make an effort to initiate participation - but instead were thrown into it.

5.2.5.3. Surfacing assumptions through decision making and discussion

Surfacing was effectively supported by the simulation. Many participants raised the fact that the feedback from the simulation partly contradicted their expectations. Particularly, the payoffs for high mitigation spending were described as surprisingly low. Some of the comments explicitly referred to prior course lectures as the cause for these expectations. Other comments (e.g. *"it's a climate game, the answer has to be..."*) pointed more towards prior assumptions that participants had brought to the course. These apparently included common stereotypes about climate education.

The simulation's effectiveness in surfacing these assumptions was much appreciated by the course organisers, as it resonated well with their overall teaching philosophy. Evidently, the group's realisation that the game contradicted their personal values gave them something relevant to argue about. Framing their arguments required peers to give explicit, verbal form to those sentiments and exposed them to reflection and counterarguments.

5.2.5.4. Subversive play is O.K.

Despite knowing that the goal was to "make money", some regions intentionally played towards lower temperature, justifying their decisions with a narrative of "saving the planet". Unsurprisingly, this conflict caused some friction within teams, since participants knew that they had to collaborate in order to win. Surprisingly, however, rather than resulting in classroom-wide chaos and a burden to the teacher, conflict was generally sorted out within teams on a peer level, using a mix of argumentation, joking and laughter.

The concept that players may intentionally substitute the game's proposed values with self-constructed goals is an emergent theme in the gaming literature (Salen & Zimmerman 2004). Various notions, such as subversive play (Flanagan 2009), soft modding (Hayes & Gee, 2010) and oppositional play (Barr et al. 2005), have highlighted the potential for players to explore - systematically or opportunistically - the values that a given game implicitly promotes. In a video game context, this may involve acts of deliberate self-obstruction or even self-destruction, such as making one's in-game character run into walls or worse. This may effectively prevent players' progress in the game and impede the implicit goal of achieving mastery.

Remarkably however, in the case of 4Decades, subversive play did not defeat the purpose of the game, nor did it "subvert" the learning goals. To the contrary, arguably subversive play supported surfacing, reflection and a critical stance towards data sources. In terms of (Kratwohl 2002) the findings suggest higher-level learning benefits, especially in the metacognitive knowledge, as participants created, analysed and evaluated alternative goals.

5.2.5.5. Making subversive play feasible at classroom scale

Arguably, being able to leverage subversive play without impeding the overall activity is a huge benefit. As far as the analysis showed, this benefit was enabled by two key design factors: (a) linearity, (b) performance-independent pace and (c) decoupled team play.

A) Linearity

One key design factor that enabled this benefit is described as "linearity" (Dillenbourg and Jermann 2010), requiring the design to be *"a simple sequence of activities that almost all students will perform at almost the same period."*

B) Performance-independent pace

Although linearity guarantees that a group stays together, it does not, per se, prevent individuals from slowing a whole group down. This highlights the importance of "performance-independent pace" (from now on called PIP), by which I mean that the pace of progression through a series of tasks should be independent of how groups or individuals perform within each task.

In the video game literature, the combination of linearity and PIP is known as gaming "on-a-rail" (Juul 2002). This mechanism is found, for instance, in some arcade shooting

games where multiple players share a screen. The game automatically moves all player characters through a level together at a constant speed, preventing any player from falling behind. Similarly, 4Decades allowed the teacher to move the group forwards as a whole, independent of the decisions they made on the way. No matter how subversively a team or region played, they were never able to obstruct the overall progress, since the teacher had ultimate control over when to start and end the rounds.

PIP is somewhat supported, albeit not explicitly pronounced, in Dillenbourg & Jermann's (2010) general argument for the teacher's control and flexibility. In the specific context of large-group, task-based conflict resolution, I would strongly argue for PIP as a distinct and essential design principle. Without PIP, individuals would have the ability to delay the process intentionally. This affordance could easily be exploited by disagreeing individuals as a nasty means of protest, circumventing the need to make convincing arguments. This would clearly defeat the purpose of the activity and create a risk of unnecessary nuisance to other participants and the teacher. The importance of PIP for constraining conflict resolution cannot be emphasised clearly and strongly enough.

While it makes sense for the teacher to *dictate* the pace (i.e., to decide when to switch), the burden of *enforcing* the pace (i.e., to make sure that the switching actually happens) should be supported by the logic of the application. Without this support, teachers may potentially find themselves competing with the affordances of the technology for participants' attention. This is a tough competition, which has in the past led teachers to awkward solutions, such as disallowing tool use - which may involve physically taking away bits of technology, such as the stylus of a TabletPC (Prieto 2012) - in order to regain control. Such fighting should not be necessary in the age of tablets, which can be centrally controlled via WIFI, e.g., to control the pace.

Therefore it makes sense, with regard to pace, to distinguish between dictating and enforcing. This distinction is less pronounced in the traditional classroom (where the teacher does both) or in the above example of the arcade shooter (where the software does both). However, for the digital classroom it is useful to separate these concerns, allowing the teacher and technology to work more smoothly together.

C) Decoupled team play

Decisions made in one team should not influence the other team. Although 4Decades allowed participants to create conflict by sabotaging their own team's outcomes, this conflict was limited to the individual team level, i.e., stayed within one single table, where face-to-face mechanisms were available as appropriate means for conflict resolution.

The latter was shown to trigger productive discussions, much in line with prior research (Doise et al. 1984; Dillenbourg 2002).

Arguably, face-to-face mechanisms - such as body language and being able to identify the “saboteurs” - played an important role in this process of conflict resolution. If the teams were not decoupled, then conflicts could potentially extend to the whole room, spanning multiple tables. This would be undesirable, since it would certainly increase the risk of conflict resolution becoming more difficult, complicated, potentially awkward and potentially burdening for the teacher, without adding any apparent benefit.

In summary, all of the above three design factors - (a) linearity, (b) performance-independent pace and (c) decoupled team play - are arguably critical for leveraging the potential of subversive play as a benefit, rather than a threat, at classroom scale.

5.2.5.6. How can participants discuss modelling decisions?

Although the debriefing allowed participants to raise their concerns about the model outputs - and the teacher was able to respond with empathy and partial clarification - the debriefing was more like a Q&A with the teacher, rather than peer discussion.

There was an apparent mismatch of language palpable in this Q&A. Apparently, the vocabulary that we (as designers) had used to describe the components and relations in the climate model was not shared by participants. Instead, several comments were framed in terms of “penalty”, “punishment” and “reward”. I find this particular choice of words somewhat concerning, because it is evocative of a behaviouristic perspective that is clearly in contrast to the pedagogical intention of the design. Climate change education is not puppy training. Our intention had been to create an infrastructure for helping groups engage with the complexity and uncertainty of the topic and discuss possible future scenarios. The intention was clearly *not* to present any predefined notions of desirable behaviour, let alone reinforce them by giving rewards or punishment. That would be extremely presumptuous for a game on this topic for a mature audience.

Whether any of the participants truly subscribed to the above behaviouristic interpretation is not entirely clear from the collected data. It may well be that the words “punishment” etc. were used simply for lack of more precise vocabulary. Such a lack was evident in the debriefing discussion. The group clearly lacked a common language for discussing the essential modelling decisions, making it difficult for the teacher to help them dissect the issue in the debriefing.

This raises the questions (a) what such a common language could be, and (b) how it could be effectively integrated with the role-playing activity. Paper-based diagrams and tables could offer a partial solution, by making some of the modelling decisions more transparent. However, to really discuss alternative decisions and explore their ramifications, an interactive approach was needed, allowing participants to make deep changes to the model. Considering the group's large size and mixed skills, none of the representations that the model was originally built in were appropriate (i.e., mathematical equations, Excel and Java code). Instead, a new kind of representation had to be devised, as addressed in the next chapter.

5.2.5.7. Orchestration load

The above analysis revealed much insight into the teacher's orchestration load in the first session. A description of the teacher's responsibilities was presented, detailing what the teacher did over the course of the session, when, where, how and why; and how these actions were supported by the technology. Evidently, the teacher was able to handle all these responsibilities on his own, without being overwhelmed at any point. This was an important criterion for the design's overall feasibility.

Further tasks were added to the teacher's list of responsibilities, as the subsequent design iteration added a new kind of group task to the activity (rule changing, see next chapter). A complete list of the teacher's actions, including these additions, is presented in Appendix A as part of the CLOVER framework.

5.3. Chapter summary

The above analysis examined the overall feasibility of the 4Decades approach in a real training course. In addition to the 9 Gameplay Objectives, which were essentially met, another set of 3 design factors were identified as critical. These were: Linearity, performance-independent pace (PIP), and decoupled team play. To manage PIP appropriately, it was argued, the teacher must be in control of the pace and the application must support this control. This worked consistently well throughout the session, except for one instance (the meander vignette) where the teacher closed the round based on the information that the team had finished their decision-making, while one region was still busy. This minor incident illustrated the aspect of awareness within the Panoptica

round-table layout (cf Chapter 4), showing on the one hand the layout's fallibility and on the other hand its subtle ability to self regulate.

Although the design could reasonably be considered "good enough" for continued use, potential for further improvement was indicated by an emergent, additional learning need, namely for participants to modify the simulation rules. This need indicated further design challenges, as addressed and discussed in the next chapter.

Chapter 6.

Analysis 2: Changing simulation rules

This chapter describes a substantial design iteration, whereby the activity was extended by an additional task, allowing groups to change the simulation rules collaboratively. This iteration was motivated by two factors: (a) participants' observed lack of appropriate vocabulary for discussing modelling decisions, and (b) participants' direct suggestions that a rule-changing feature would be beneficial.

Changes to the design involved a series of tablet interfaces, minor revisions of the teams' ambient display, and an additional ambient display to show the rule definitions for each match.

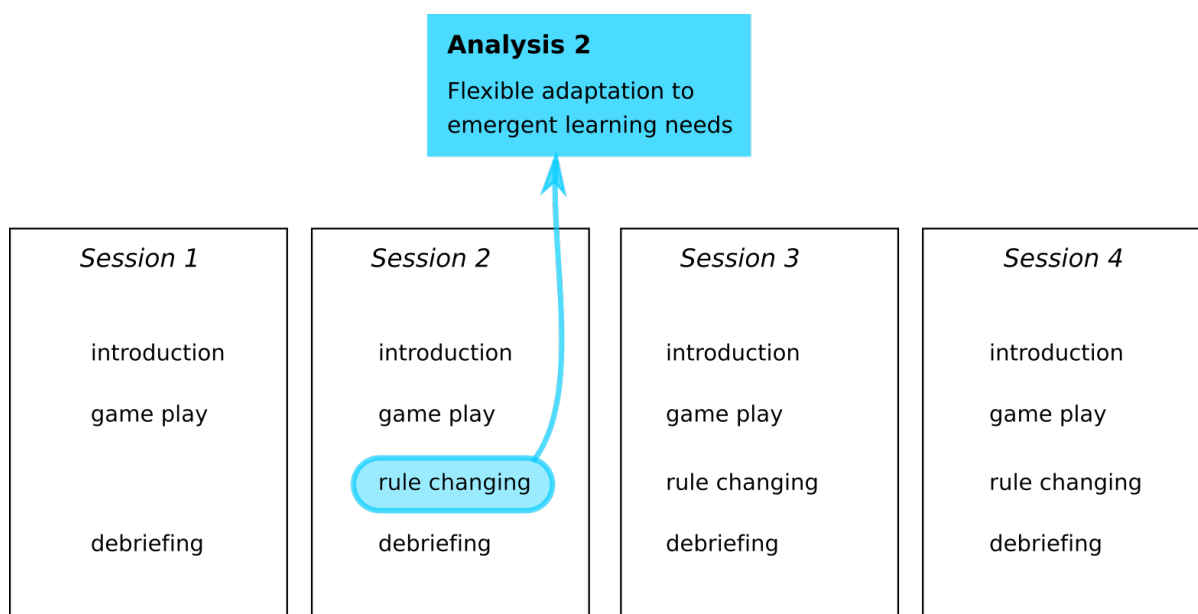


Figure 6.1.: The following analysis focuses on session 2, specifically, the rule changing

6.1. Question and aims

The question that guided this analysis was:

Q2-B. How can groups make informed changes to the simulation rules?

The aims of this analysis are:

- to identify appropriate representations and design a suitable interface to allow participants to collaborate on alternative rules for the simulation;
- to test the solution in the lab until considered mature;
- to evaluate the solution in the wild and describe the orchestration load.

6.2. Benefits of Rule-Changing

The intention behind the rule-changing feature was to encourage constructive reasoning about simulation parameters and assumptions. Arguably, the first in-situ session had somewhat fallen short in this regard, as those assumptions were treated only superficially and indirectly. Critique was mostly framed in the form of complaints, addressing the teacher and concerning the resulting feedback (black box), rather than the underlying rules of the model (glass box).

Given the outcomes of the literature review, it was conjectured that allowing groups to collaboratively change the rules might transform the activity in the following beneficial ways:

1. Participants reason collaboratively, rather than just individually.
2. Critique is framed in discussion, rather than complaints.
3. Peers are used as a resource in addition to the teacher.
4. Reasoning focuses on reflection, rather than punishment and reward.
5. Rules are explored, rather than received.
6. Groups immerse themselves more deeply in the activity.
7. Focus shifts from pure gameplay to aspects of modelling and game design.

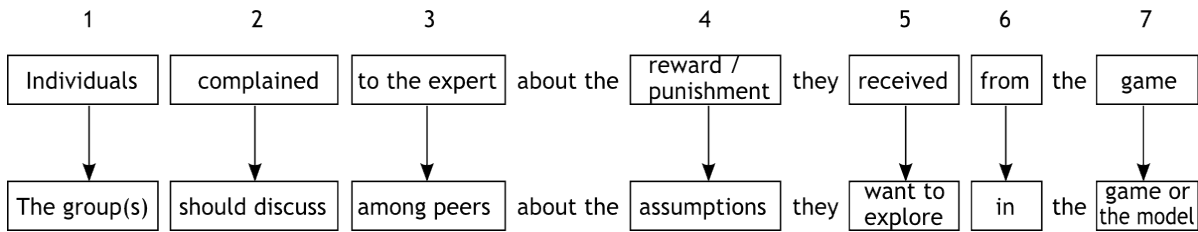


Figure 6.2.: Seven benefits of Rule-Changing

The above list was used in the following analysis to evaluate the instructional value of collaborative rule changing.

Another, more compact representation of the above list is attempted in Figure 6.3. Two sentences are presented in parallel, in order to illustrate the proposed transformation as *before and after*. Thus, both sentences touch on each of the 7 aspects, with the lower one representing a more desirable outcome. With regard to the data used in the analysis, the upper sentence represents concrete findings from session 1, whereas session 2 (with the rule changing) aimed to realise the ideals described in the lower sentence.

6.3. Design iteration

To make rule-changing feasible and effective, the aim was for the group to alternate between playing the game and changing the rules. Accordingly, we must now distinguish between two types of decision tasks: in-game decisions and game-changing decisions. Both tasks were intended to encourage discussion in teams, although via different mechanisms and for different purposes.

As detailed below, the new design challenge involved providing groups with relevant choices of alternative climate scenarios, devising adequate representations and implement the solution in an “orchestrable” way. As before, designing for orchestrability implied respecting the teacher’s real-time orchestration load, time and other constraints. Furthermore, the solution should also avoid introducing excessive amounts of additional technology.

6.4. Design objectives: adapted from Gameplay

To guide the design of the rule-changing feature and its integration in the overall activity, some of the 9 Gameplay Objectives (that had previously guided the design of gameplay) were adapted and reused.

- Allow participants to apply relevant course knowledge in practice by discussing alternative rules and scenarios (adapted from Objective 1);
- Ensure that the interfaces take minimal effort to learn and use (adapted from Objective 2);
- Ensure that a single teacher is able to facilitate the rule-changing (adapted from Objective 3);
- Encourage participants to reflect on assumptions through decision making and discussion (Objective 4);
- Leverage existing team structures to facilitate collaboration and give both teams equal opportunities (adapted from Objective 7);
- Use ambient information to enable whole-group awareness of the rules and reflection on previous rules (adapted from Objective 8);
- Allow all participants to contribute equitably to the rule-changing and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics (adapted from Objective 9).

Keeping the above objectives and the Benefits of Rule-Changing in mind, the design set out to (a) provide groups with relevant choices of alternative climate scenarios and winning conditions, (b) devise appropriate representations and (c) give participants a sense of the model author's perspective.

6.4.1. Providing relevant choices of scenarios and winning conditions

The first version of 4Decades provided one single climate scenario that had been hard-coded in the application, based on various predictions that were selected (at design time, by the model designer!) from relevant domain literature. In the new version, the aim

was to encapsulate some of this selection process and make it accessible to participants during the application's runtime.

More specifically, the selections concerned (a) climate/economic scenarios and (b) winning conditions. Whereas (a) pertains to the modelling, (b) is a question of game design, according to the definition in Chapter 1. Figure 6.3 recapitulates this definition and adds the suggestion of “pluggable” scenarios and winning conditions.

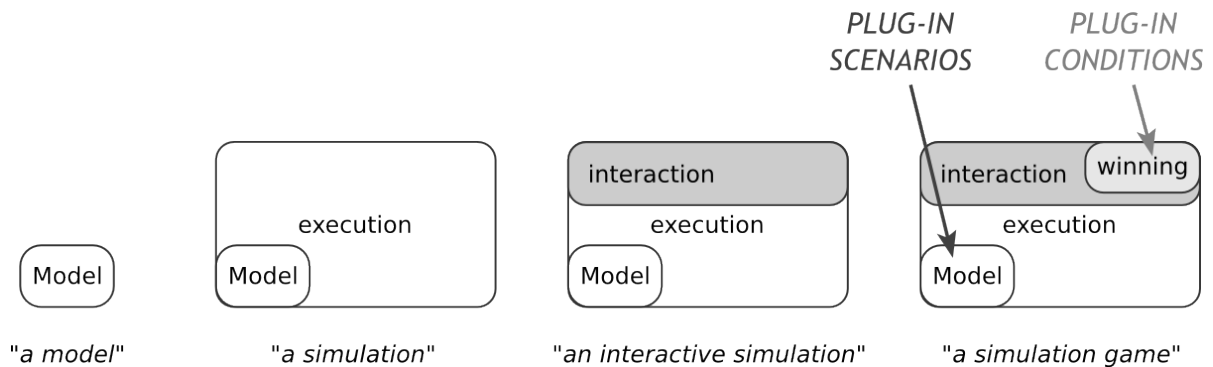


Figure 6.3.: Scenarios and winning conditions can be flexibly integrated as plugins at runtime

Exchanging these two components makes it possible to explore alternative scenarios. For instance, what if future developments made CO₂ mitigation extremely cheap (or expensive, respectively) - how would that impact the payoffs and therefore the best strategies? What if adaptation was more/less effective? Are we collaborating as a whole planet or rather competing regions? These and similar assumptions could be expressed by tweaking (a) the winning condition and (b) the cost/efficiency functions for mitigation and adaptation.

(A) winning condition. To address the need for different winning conditions, we decided to let participants choose between any of the five existing variables in the model that were considered potentially meaningful: income, emissions, temperature, baseline damage and residual damage. Further options were whether to play for high or low values and whether winning should be per planet or per region.

(B) cost and efficiency functions. The cost of mitigation and the efficiency of adaptation were represented as mathematical functions in our model. They could be represented, e.g., as equations, curves or tables.

6.4.2. Devising adequate representations

Which representation was most suitable: equations, curves or tables? Should participants draw curves by hand? Deform curves via breakpoints? Type in numbers? Write mathematical formulae? Again, the interfaces had to be extremely quick and easy for everyone to understand, whilst also being transparent, shareable and easy to use. Several options were explored through sketching. The final design favoured simple text and numbers.

6.4.3. Giving participants a sense of the model author's perspective

Creating scenarios for a simulation involves dealing critically with modelling decisions. This had been evident when designing the original equations behind 4Decades. Since the sources relied on different models whose predictions of the global economy varied substantially, choosing “the right” values and functions for the simulation had been a profound challenge. It had involved critical assessment of alternative assumptions and their implications to the predicted scenarios. Moreover, the challenge of abstracting a complex topic into something simple enough that one can *play* - without unduly veiling the essential complexity - was almost paradoxical. Stephen, being a domain expert, repeatedly emphasised how much he had gained from this exercise. Based on this experience, it was conjectured that participants might benefit in similar ways, given an accessible design.

6.4.4. Reusing tablets and ambient displays

Could the existing tablets be re-used as input devices for the game-changing? To avoid excessive technical requirements and excessive learning time, we were hesitant to introduce any further equipment unless needed. Since eight tablets were already part of the design, our first approach was to consider whether (and if so, how) up to eight tablets could allow 20-30 participants to make rule-changing decisions together.

A number of sketches were made, involving different numbers of tables for different sizes of subgroups. In the end, the simplest and least contrived option seemed to be a single-tablet interface that is shared within a team of 10-15 participants.

6.5. Changes to the application

The following new features were introduced.

- Tablet interfaces for rule changing
- An additional ambient display for the rules
- Minor changes to the team display, including highlights to help participants analyse the data in relation to the dynamic winning conditions.

Below I describe how these features were designed and integrated into the overall activity.

6.5.1. Tablet interfaces for rule changing

Several iterations led to interface prototypes which were

- **quick and easy to understand**, due to intuitive arrangements of simple sentences and buttons;
- **easy to read from a distance**, due to large, static text;
- **easy to share**, by avoiding scrolling and complex touch interactions.

Radio buttons with check symbols allowed users to select one of several options. The buttons were designed in the simple round and white design that was familiar from the gameplay interfaces, so that participants would immediately know where to tap.

Below I describe the tablet interfaces that made it into the final design, after some user testing in the lab. Versions that failed the lab tests are discussed further below in Section 6.8.

Figure 6.4 shows the final tablet interface for selecting the winning condition. The whole screen can be read as one concise sentence. The goal was to keep the rules very simple, clear and memorable. Possible choices also include counter-intuitive settings, such as, “the region with the highest damage”, enabling groups to take a reverse problem solving approach if they wished.

Figure 6.5 shows the tablet interface for selecting the mitigation scenario. A single selection menu was chosen in response to the lab tests which had indicated that too many options can confuse participants and delay the process. Therefore, relevant sets of numerical values were bundled into “scenarios”. Three scenarios were provided to

GAME - Winning Condition

The winner will be

the ☒ planet with the ☐ highest ☐ income
☐ region ☒ lowest ☒ emissions
☐ temperature
☐ bas. damage
☐ res. damage

at the end of the match.

Figure 6.4.: Tablet interface for selecting the winning condition.

represent more optimistic and pessimistic views on climate change, in addition to the default scenario. To give the scenarios useful names, descriptive labels were chosen in accordance with relevant literature that participants were familiar with: “techno optimist”, “business as usual” and “fossil fuel abundance”. The numbers in the bottom half show the contents of the currently selected scenario.

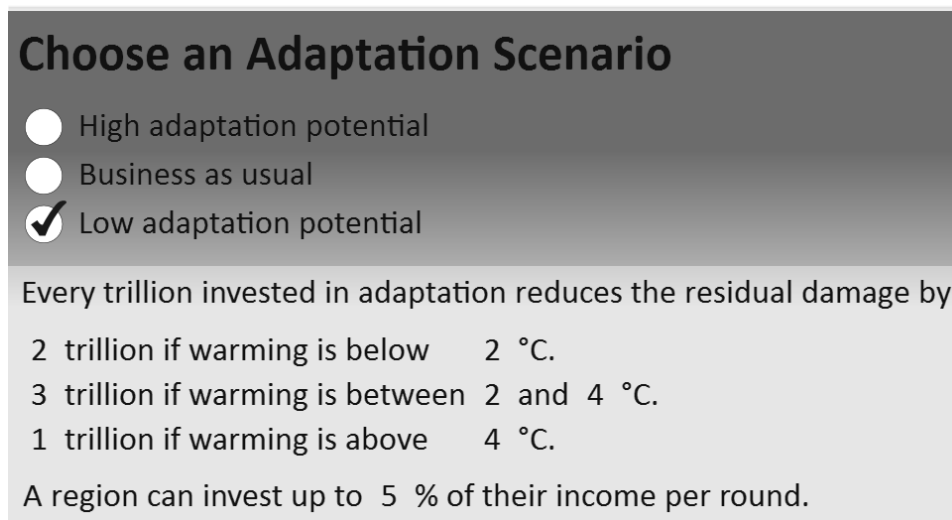
Choose a Mitigation Scenario

☐ Techno optimist
☐ Business as usual
☒ Fossil fuel abundance

| | | |
|-------------------------------------|----|-------------|
| the 1st gigatonne in a round costs | 1 | \$trillion. |
| the 2nd gigatonne in a round costs | 2 | \$trillion. |
| the 3rd gigatonne in a round costs | 4 | \$trillion. |
| the 4th gigatonne in a round costs | 8 | \$trillion. |
| the 5th gigatonne in a round costs | 16 | \$trillion. |
| the 6th gigatonne in a round costs | 32 | \$trillion. |
| the 7th, 8th, 9th, etc... each cost | 64 | \$trillion. |

Figure 6.5.: Tablet interface for selecting the mitigation scenario

The adaptation scenarios were constructed after the same principle (Figure 6.6). Each of three scenarios entails values for payoffs (left column), temperature thresholds (right) and an upper limit on the amount of adaptation (bottom row). The latter feature was newly introduced in this version of the application, to prevent teams from buying themselves out of global warming too cheaply via adaptation.



Choose an Adaptation Scenario

☐ High adaptation potential
☐ Business as usual
☒ Low adaptation potential

Every trillion invested in adaptation reduces the residual damage by

- 2 trillion if warming is below 2 °C.
- 3 trillion if warming is between 2 and 4 °C.
- 1 trillion if warming is above 4 °C.

A region can invest up to 5 % of their income per round.

Figure 6.6.: Tablet interface for selecting the adaptation scenario

The above three menu pages were designed to be presented in this sequence. To progress from one page to the next, the same mechanism was used as in the gameplay, i.e., putting the teacher in charge of deciding when to switch. Upon the teacher’s verbal cue (e.g. “Are you done deciding? O.K. next page”), a researcher (myself) pressed a button on the control laptop, causing the tablet to switch to the next page.

6.5.2. An additional ambient display for the rules

Since participants could not be expected to memorise all the different rules for multiple matches, an additional display was needed to keep a record of the winning condition, mitigation scenario and adaptation scenario for each match (Figure 6.7). The design emphasises easy legibility of key data and maintains the same vertical match-by-match structure as the team displays.

| 4Decades GAME SETTINGS | | | | | | | | | |
|---|--|--------------------------------------|----|----|---|---|-----------------------|-----|----|
| MATCH 1 | | Mitigation cost | | | | | Adaptation efficiency | | |
| The winner is the PLANET with the HIGHEST INCOME. | | GtCO ₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | \$trillion | -1 | 1 | 2 | 4 | 8 | 16 | 32 |
| | | | | | | | below 2°C | x 3 | |
| | | | | | | | 2 - 4°C | x 6 | |
| | | | | | | | above 4°C | x 2 | |
| | | Regions can invest up to 10 % of GRI | | | | | | | |
| MATCH 2 | | Mitigation cost | | | | | Adaptation efficiency | | |
| The winner is the PLANET with the LOWEST TEMPERATURE. | | GtCO ₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | \$trillion | -1 | 1 | 2 | 4 | 8 | 16 | 32 |
| | | | | | | | below 2°C | x 3 | |
| | | | | | | | 2 - 4°C | x 6 | |
| | | | | | | | above 4°C | x 2 | |
| | | Regions can invest up to 10 % of GRI | | | | | | | |
| MATCH 3 | | Mitigation cost | | | | | Adaptation efficiency | | |
| The winner is the PLANET with the HIGHEST INCOME. | | GtCO ₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | \$trillion | -1 | 1 | 2 | 4 | 8 | 16 | 32 |
| | | | | | | | below 2°C | x 3 | |
| | | | | | | | 2 - 4°C | x 6 | |
| | | | | | | | above 4°C | x 2 | |
| | | Regions can invest up to 10 % of GRI | | | | | | | |
| MATCH 4 | | Mitigation cost | | | | | Adaptation efficiency | | |
| The winner is the PLANET with the HIGHEST INCOME. | | GtCO ₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | \$trillion | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
| | | | | | | | below 2°C | x 2 | |
| | | | | | | | 2 - 4°C | x 3 | |
| | | | | | | | above 4°C | x 1 | |
| | | Regions can invest up to 5 % of GRI | | | | | | | |
| MATCH 5 | | Mitigation cost | | | | | Adaptation efficiency | | |
| The winner is the PLANET with the HIGHEST INCOME. | | GtCO ₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | \$trillion | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
| | | | | | | | below 2°C | x 4 | |
| | | | | | | | 2 - 4°C | x 8 | |
| | | | | | | | above 4°C | x 4 | |
| | | Regions can invest up to 20 % of GRI | | | | | | | |

Figure 6.7.: The additional display for the rules

6.5.3. Changes to the team display

The team displays were slightly changed (Figure 6.8). For instance, a gentle colour tone was added. The fonts were slightly more stylish. To make space for a potential fifth match, the text font was made flatter and the lines that said “Match 1”, “Match 2”, etc., were removed in favour of a large number on the left. The numbers for income were rounded to integers, in order to make that whole section faster to read. Rounding affected only the display, but not the values in the model. A “SUM” column was added on the far right to show the sum of all regions’ results for each round. This replaced the “Final GPI” (Gross Planetary Income) label in the previous version and made it easier to see which team was in the lead in any particular round. Finally, yellow highlights were introduced to highlight the winning criterion and to indicate which team was in the lead.

Alpha Centauri

| | Mitigation | | | | Adaptation | | | | Emissions | | | | °C | Baseline Damage | | | | | Residual Damage | | | | GRI | | | | SUM |
|---|------------|---|---|---|------------|---|---|---|-----------|-----|-----|-----|-----|-----------------|-----|-----|-----|-----|-----------------|-----|-----|----|-----|----|----|-----|-----|
| 1 | 3 | 4 | 4 | 2 | 1 | 1 | 1 | 1 | 6.2 | 5.2 | 5.2 | 7.2 | 1.1 | 1.4 | 1.1 | 1.1 | 1.5 | 0.0 | 0.0 | 0.0 | 0.0 | 21 | 17 | 17 | 23 | 82 | |
| | 3 | 3 | 4 | 3 | 2 | 1 | 1 | 2 | 4.5 | 3.5 | 2.5 | 5.5 | 1.8 | 2.9 | 2.3 | 1.9 | 3.2 | 0.0 | 0.0 | 0.0 | 0.0 | 25 | 20 | 16 | 28 | 91 | |
| | 1 | 2 | 3 | 2 | 3 | 2 | 2 | 3 | 4.7 | 2.7 | 0.7 | 4.7 | 2.4 | 5.0 | 4.0 | 2.9 | 5.4 | 0.0 | 0.0 | 0.0 | 0.0 | 33 | 26 | 19 | 35 | 114 | |
| | 1 | 0 | 1 | 1 | 3 | 1 | 2 | 3 | 5.0 | 4.0 | 1.0 | 5.0 | 3.0 | 8.0 | 6.4 | 4.7 | 8.7 | 0.0 | 0.4 | 0.0 | 0.0 | 43 | 35 | 25 | 47 | 152 | |

| | Mitigation | | | | Adaptation | | | | Emissions | | | | °C | Baseline Damage | | | | | Residual Damage | | | | GRI | | | | SUM |
|---|------------|---|---|---|------------|---|---|---|-----------|------|------|------|-----|-----------------|-----|-----|-----|-----|-----------------|-----|-----|---|-----|---|---|--|-----|
| 2 | 5 | 5 | 5 | 5 | 0 | 0 | 0 | 0 | 4.2 | 4.2 | 4.2 | 4.2 | 0.9 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 9 | 9 | 9 | 9 | | |
| | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 0 | 1.5 | 1.5 | 1.5 | 1.5 | 1.3 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 6 | 6 | 6 | 6 | | |
| | 3 | 3 | 3 | 3 | 0 | 0 | 0 | 0 | -0.2 | -0.2 | -0.2 | -0.2 | 1.4 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 6 | 6 | 6 | 6 | | |
| | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 0 | -3.0 | -3.0 | -3.0 | -3.0 | 1.1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 3 | 3 | 3 | 3 | | |

| | Mitigation | | | | Adaptation | | | | Emissions | | | | °C | Baseline Damage | | | | | Residual Damage | | | | GRI | | | | SUM |
|---|------------|---|---|---|------------|---|---|---|-----------|-----|-----|-----|-----|-----------------|-----|-----|-----|-----|-----------------|-----|-----|----|-----|----|----|-----|-----|
| 3 | 4 | 3 | 4 | 4 | 1 | 1 | 1 | 1 | 5.2 | 6.2 | 5.2 | 5.2 | 1.0 | 1.1 | 1.3 | 1.1 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 17 | 21 | 17 | 17 | 72 | |
| | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 4.5 | 5.5 | 4.5 | 4.5 | 1.8 | 2.5 | 3.1 | 2.5 | 2.5 | 0.0 | 0.1 | 0.0 | 0.0 | 22 | 28 | 22 | 22 | 96 | |
| | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 3.7 | 4.7 | 3.7 | 3.7 | 2.5 | 4.6 | 5.7 | 4.6 | 4.6 | 0.0 | 0.0 | 0.0 | 0.0 | 30 | 38 | 30 | 30 | 129 | |
| | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2.0 | 3.0 | 2.0 | 3.0 | 3.0 | 7.0 | 8.8 | 7.0 | 7.4 | 0.0 | 0.0 | 1.0 | 0.0 | 38 | 48 | 38 | 40 | 165 | |

| | Mitigation | | | | Adaptation | | | | Emissions | | | | °C | Baseline Damage | | | | | Residual Damage | | | | GRI | | | | SUM |
|---|------------|---|---|---|------------|---|---|---|-----------|-----|-----|-----|-----|-----------------|------|------|------|-----|-----------------|-----|-----|----|-----|----|----|-----|-----|
| 4 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 1 | 6.2 | 6.2 | 6.2 | 6.2 | 1.1 | 1.7 | 1.7 | 1.7 | 1.7 | 0.0 | 0.0 | 0.0 | 0.0 | 26 | 26 | 26 | 26 | 104 | |
| | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 4.5 | 4.5 | 4.5 | 4.5 | 1.9 | 4.4 | 4.4 | 4.4 | 4.4 | 0.4 | 0.4 | 0.4 | 0.4 | 36 | 36 | 36 | 36 | 147 | |
| | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2.7 | 2.7 | 2.7 | 2.7 | 2.5 | 7.8 | 7.8 | 7.8 | 7.8 | 1.8 | 1.8 | 1.8 | 1.8 | 50 | 50 | 50 | 50 | 200 | |
| | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1.0 | 1.0 | 1.0 | 1.0 | 2.8 | 11.6 | 11.6 | 11.6 | 11.6 | 2.6 | 2.6 | 2.6 | 2.6 | 66 | 66 | 66 | 66 | 266 | |

| | Mitigation | | | | Adaptation | | | | Emissions | | | | °C | Baseline Damage | | | | | Residual Damage | | | | GRI | | | | SUM |
|---|------------|---|---|---|------------|---|---|---|-----------|------|------|------|-----|-----------------|------|------|------|-----|-----------------|-----|-----|----|-----|----|----|-----|-----|
| 5 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 9.2 | 9.2 | 9.2 | 9.2 | 1.3 | 1.9 | 1.9 | 1.9 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 23 | 23 | 23 | 23 | 92 | |
| | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 10.5 | 10.5 | 10.5 | 10.5 | 2.9 | 5.4 | 5.4 | 5.4 | 5.4 | 0.0 | 0.0 | 0.0 | 0.0 | 31 | 31 | 31 | 31 | 124 | |
| | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 3 | 11.7 | 11.7 | 11.7 | 11.7 | 4.7 | 11.7 | 11.7 | 11.7 | 11.7 | 0.0 | 0.0 | 0.0 | 0.0 | 40 | 40 | 40 | 40 | 160 | |
| | 0 | 0 | 0 | 0 | 6 | 6 | 6 | 6 | 13.0 | 13.0 | 13.0 | 13.0 | 6.7 | 21.4 | 21.4 | 21.4 | 21.4 | 0.0 | 0.0 | 0.0 | 0.0 | 49 | 49 | 49 | 49 | 197 | |

Figure 6.8.: The team display in session 2 (showing 2 out of 5 matches)

6.6. Integrating rule-changing into the whole activity

Rule changing was realised as a temporally separate process, i.e., between the matches. This decision was motivated by several considerations. From a gaming point of view, the alternative (i.e. changing the rules while playing) would be confusing and difficult to orchestrate. Pedagogically speaking, the idea of alternating between gameplay and rule-changing promises the benefit that groups can change the rules in immediate response to emergent questions that triggered their curiosity during gameplay. From a practical perspective, the same input devices can be used for both modes, if they happen at different times. With regard to methodology, separating the two modes makes them easier to evaluate individually.

6.7. Lab testing

One large-group lab trial was conducted to test the usability of the rule-changing feature, to verify that the application was still stable after implementing the new features, and to rehearse the new procedures. As with previous lab tests, participants were adults, recruited from the university who took part for fun and free snacks. The activity was tested as a whole, allowing the teacher and myself to practice all the procedures under somewhat realistic conditions. This included guiding participants through the rule-changing procedures and letting them fill in a paper-based questionnaire that had been designed for the real training course.

As a result of the lab test, minor technical bugs were fixed and the questionnaire was slightly revised. Moreover, early versions of the rule-changing interface were rejected or improved, as detailed below.

6.8. Rejected ideas for the tablet interface

As a result of the lab testing, some of the suggested interface prototype were ruled out, as they were found to lack some relevant use qualities. Since these use qualities are worth pointing out in the context of this research, the rejected prototypes are presented below.

Figure 6.9 shows an early design for the adaptation menu. Hundreds of input combinations were possible by changing each number separately. In the lab test this version took far

too much time, as the team ended up discussing each number, while the other team was waiting impatiently. Consequently, the revised design offered only a small number of choices, that could be made quickly.

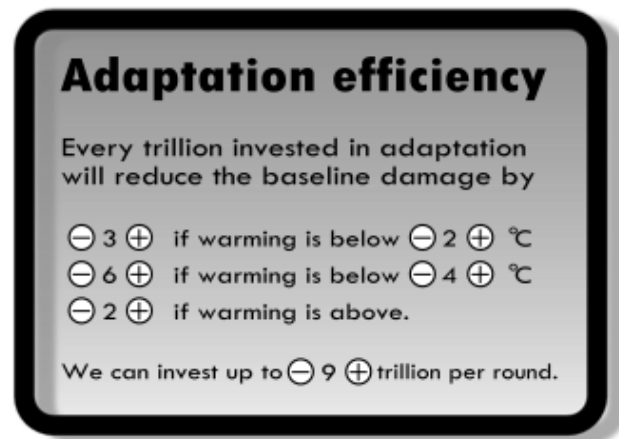


Figure 6.9.: Too many choices: Rejected version of the interface for selecting adaptation settings

Figure 6.10 shows a rejected interface for showing or hiding the forecast. The idea was inspired by a participant's suggestion, with the aim to create suspense by increasing uncertainty. However, since the idea was only vaguely related to the core topic, the teacher expressed doubts, arguing "I wouldn't know how to debrief this."

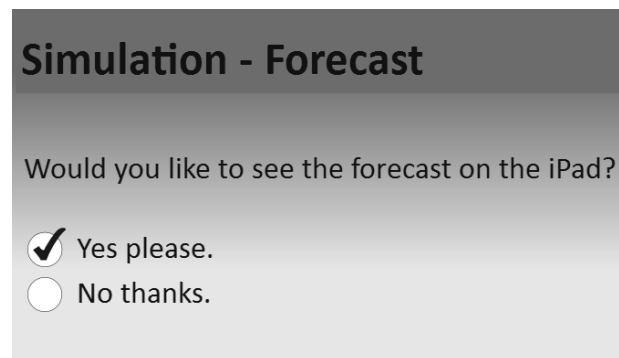


Figure 6.10.: Slightly off-topic: Rejected interface for showing or hiding the forecast

Figure 6.11 shows another option that was rejected. Allowing participants to think about appropriate time scales and granularity initially seemed to be a compelling feature. However, on second thought, two major problems with this idea became apparent. Firstly, if the number of rounds was flexible, then how could the teacher ensure that the whole activity stayed within schedule? Secondly, if the number of rounds was flexible, how could we avoid running out of vertical space on the ambient display?



Figure 6.11.: Breaking control: Rejected interface for selecting the number of rounds and years per round

6.9. Evaluation in the wild

After the lab test and some subsequent revisions, the new version of 4Decades was evaluated in a real training course. The goal of the evaluation was to test the usability of the rule-changing feature and explore the extent to which participants and the teacher find it useful. Importantly, the goal was *not* to make any comparison between the results of this session and the previous session, such as comparing group behaviour, learning effects or the quality of discussions during or between matches. Such a comparative approach was not considered feasible, because the two sessions were conducted under different circumstances (described in Section 6.11) which would have confounded any such comparison.

6.10. Method

The same data collection methods were used as in the first session, except that a paper-based questionnaire was provided for participants to fill in on-site. The questionnaire included Likert scales and open questions regarding usability and satisfaction, with an emphasis on the newly added features. The analysis mainly relied on these questionnaires, video data, partial transcripts of group interaction and a full transcript of the debriefing.

6.11. Setting

The course was similar to the previous one with regard to group size (26) and participants' age range, gender ratio and educational background. However, the circumstances of the course differed in a number of ways:

1. The course topic was on sustainability in general, including but not limited to climate;
2. The simulation was scheduled at the beginning of the course, since the course organisers had insisted on using it as a social ice breaker. Instead of two days worth of lectures, participants only received a 20-minute introduction on the specific topic;
3. The group had never met before, whereas the previous group had already spent two days together.

6.12. Room layout and participant roles

Figure 6.12 shows the new room layout (right) with the added display in the middle. Chairs were provided for all participants this time, based on the assumption that the unacquainted group would be more comfortable if the activity treated them all equally from the beginning, rather than splitting into roles. No distinction was made between players and analysts in this session.

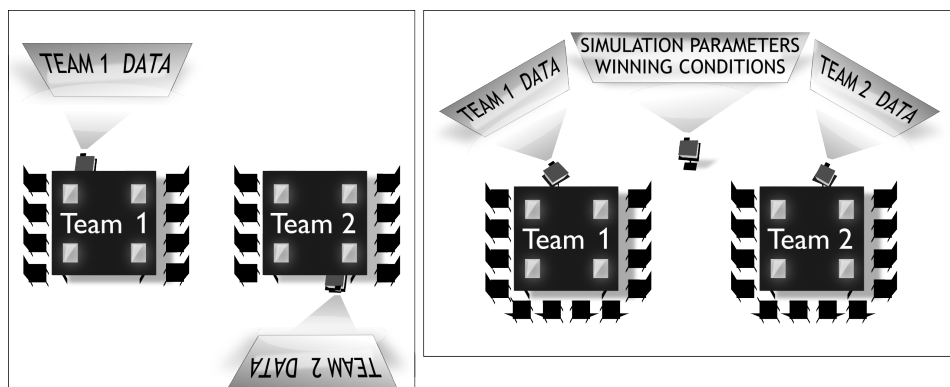


Figure 6.12.: Room layout in session 2 (right) in contrast to session 1 (left)

6.13. Procedure

After the introduction and a practice match, 5 proper matches were played. The first match was played by the original rules. Afterwards, the teams took turns defining the rules for the next match. Each team thus had two opportunities to define the rules. Figure 6.13 shows an overview of the entire 94 minutes including introduction and debriefing.

As the figure shows, the overall rhythm of the session was similar to the previous session. The introduction and practice match also lasted about 15 minutes and each proper match took between 5 and 8 minutes (average 6.4). Also shown in the figure is a 5-minute technical interruption before the third match and two breaks for participant questionnaires, totalling 9 minutes. The four rule-changing discussions each lasted between 1 minute and 2.5 minutes. The figure also includes a visual summary of the teacher's actions which are discussed further below.

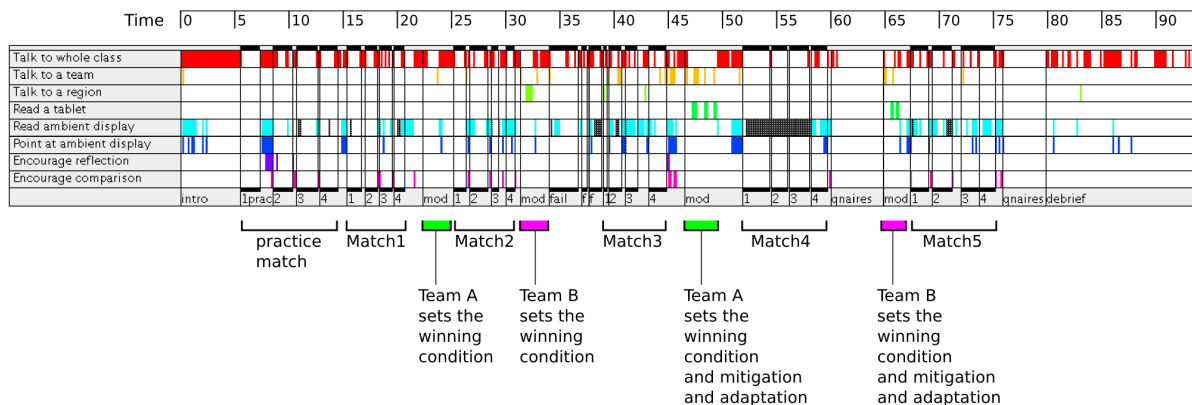


Figure 6.13.: Overview of session 2, including the rule-changing moments of Team A (green) and Team B (magenta)

6.14. Findings

The following presentation of findings is structured according to the following themes.

- Chosen rules
- Participants' reflections on the chosen rules
- Large groups sharing one tablet
- Dialogue during rule-changing

- Clarification among peers
- Participants' overall ratings of the activity
- Comments in the debriefing
- The teacher's actions
- The teacher's reflections on rule-changing

6.14.1. Chosen rules

Figure 6.1 shows an overview of the scenarios and winning conditions that the groups selected. To avoid overwhelming the groups with choices or spending too much time, only a subset of the choices were available each time. The minus signs in the figure indicate occasions where the default rules were used - i.e., the same rules as in session 1 - i.e., the winner being the planet with the highest income and “business as usual” as mitigation and adaptation scenarios. As the figure shows, the first match was played by the default rules. The second and third match allowed groups to select the winning condition. The last two matches allowed selecting the mitigation and adaptation scenarios while the winning condition was locked to default.

| Match | Team | Winning condition | Mitigation scenario | Adaptation scenario |
|-------|------|--|-----------------------|---------------------|
| 1 | - | - | - | - |
| 2 | A | Planet / lowest temperature | - | - |
| 3 | B | Planet / highest income “below 3 degrees” | - | - |
| 4 | A | - | Techno optimist | Low potential |
| 5 | B | - | Fossil fuel abundance | High potential |

Table 6.1.: The chosen scenarios and winning conditions

What these selections looked like on the ambient display was presented further above in Figure 6.8. Reflected in the groups' choices (and questionnaire comments) was a strong preference for scenarios that the group perceived as realistic. This was particularly evident in the choices of winning conditions. For instance, none of the counter-intuitive

winning conditions (e.g. “lowest income”) was chosen, although some of them could have made for an interesting game.

Different choices were made for each match. Winning was always chosen per planet (never per region).

In match 3 the team selected “highest income” and made an additional verbal agreement with the teacher that the temperature increase had to be kept below 3 degrees in order to win. The teacher was aware that the application by design did not support *combined* winning conditions. However, rather than submitting to this limitation, the teacher offered the group to play by their suggested rules and simply ignore the winning-related feedback from the simulation. The following partial transcript describes how this was negotiated. The full transcript can be found in Appendix A.

Teacher: So we could try and play that, actually, if that's what you want to do. That'll just mean that that screen will be wrong for match 3. So, sure, we could do that. Do you wanna do that? Many: yes! [...] Teacher: You're gonna have to help me work out the winner. You'll see at the end of this - if we do it your way, which I agree is a very interesting way, we'll have to make out the winner ourselves. [...] Teacher: OK. Right. So just put anything in that iPad. [...] Teacher [to the whole group]: OK Betelgeuse has chosen their winning condition, which on the screen is written like this - ignore that, it's rubbish, right? So, we'll do it this way. They want to play a bit more sophisticated game, which is quite interesting. [Name anonymised], explain what the winning condition is.

6.14.2. Participants' reflections on the chosen rules

According to the groups' choices and individual questionnaire comments, the majority of participants appeared to prefer a differentiated definition of winning, e.g. high income *and* low temperature. Several commentators argued that this made the simulation more realistic and helped them empathise with different perspectives on the topic, including the situation of real-life global policy makers. Some comments furthermore argued that combined winning conditions should be included as a standard feature in future versions.

When asked in the questionnaires which match they found most interesting, the majority (18 out of 26) said the third match. Some of those comments explicitly attributed this to the combined winning condition, which was strongly defended in the debriefing by several participants:

- *Here this team won with very high temperature. And this is something that you're probably going to face in the next decades. Right? So, I mean if you're supposed to learn from this, you have to have some consequence for people who just maximise profit. This is why you can never have just one condition, either lowest temperature or highest income. You have to mix.*

6.14.3. Large groups sharing one tablet

Figure 6.14 shows 12 participants huddled together over one iPad to change the game rules. Some had left their chairs for this purpose. The other team can be seen in the background, using the break for a team discussion.



Figure 6.14.: Changing the rules: 12 participants sharing an iPad

Although reading the tablet required some individuals to bend their necks and tilt their heads, all appeared able to read the text with reasonable effort - with the exception of one participant (bottom of figure) who had stayed in her chair despite the fact that this position did not allow her to read the tablet. The data did not provide a definitive explanation for this behaviour, which occurred only once.

In the questionnaires all 26 participants responded with “yes” to the question whether they found the tablets easy to use. Unfortunately, this question had failed to differentiate explicitly between gameplay and rule changing. Judging from the video data, in general participants did not seem to mind the brief periods of minor physical discomfort. The rule changing never lasted more than 2.5 minutes.

6.14.4. Dialogue during rule-changing

While debating which rules to choose, typically there was only one person talking at a time. Although brief utterances were made spontaneously and sometimes overlapped, there was nevertheless - most of the time - one overall stream of conversation that everyone at the table could follow. This finding was consistent across all four instances of rule-changing. It was markedly different from the gameplay discussions which usually involved multiple lines of conversations simultaneously, often one per tablet.

In three out of four rule-changing instances, one participant initially read the whole screen out loud, ensuring that everyone was - quite literally - on the same page. This was supported by the design of the interface which deliberately aimed to be readable as plain English sentences.

When arguing for or against a particular choice, participants used a diverse range of rationales. Below are some example quotes, taken from transcripts of all four rule-changing discussions. The original transcripts are included in Appendix A.

1. *“Temperature is probably easier”*
2. *“Having said this, actually it might be too easy because [the optimal gameplay strategy for] the lowest temperature is just mitigation, mitigation, mitigation - you know - five, five, five.”*
3. *“I’m happy with temperature.”*
4. *“...because lowest temperature is basically what it all comes down to”*
5. *“We’re good at money.”*
6. *“Ideal conditions are [...] global temperature stays below 2 degrees.”*
7. *“Let’s try base damage and see what that does.”*
8. *“Sounds good.”*

9. *“It would be more interesting to go low. High is the sexy one but low is probably the more interesting one.”*

10. *“We haven’t played fossil fuel.”*

Typically, after a particular choice had been made (or at least proposed), suggestions were raised and discussed regarding the choice’s likely implications for gameplay strategies. The following two excerpts illustrate this.

- *“That means that we’ll do lots of adaptation. Huge lots” - “We’ll do lots of mitigation, right?” - “Well, no, adaptation”*
- *“The third gigatonne you’ll break even” [...] “So that means we can afford a lot more mitigation” [...] “O.K., so we ain’t gonna need to fix climate change, we can just adapt” [...] “You still do, but it’s cheaper now to [mitigate]”*

6.14.5. Clarification among peers

Changing the rules allowed a team to reflect together on shared understanding. The following excerpt illustrates how a team discussed the likely implications of a particular mitigation scenario (“techno optimist”, before Match 4). One participant (“B”, highlighted in bold font) played a central role in this vignette, by sharing her predictions and conclusions with the whole team, as if to check whether her understanding was the same as everyone else’s.

- A: [In this setting] we’ll break even at the second [gigatonne].
- **B: So that means we can afford a lot more mitigation.**
- Several: Yeah / That’s right / Exactly
- C: Let’s go for that one.
- **B: O.K., so we ain’t gonna need to fix climate change, we can just adapt.**
- D: Well, erm, no.
- A: You still do, but it’s cheaper now to...
- **B: invest?**
- Several: into new technologies. / into mitigation.
- A: than it was before.

- **B: Correct, correct. Mitigation is cheaper than adaptation.**
- E: Not cheaper necessarily, but cheaper than business as usual.
- Several: Yeah / Mhm / Right.

Characteristic to this discussion was that one person made suggestions to which several team members then responded spontaneously with confirmation or clarification. Although the responses sometimes overlapped, there was nevertheless a single stream of conversation that the whole team was able to follow. Indeed, this part of the discussion concluded with a correct and relevant observation about the chosen scenario, namely that mitigation was cheaper than in the other scenario (“business as usual”) but not necessarily cheaper than adaptation.

6.14.6. Participants’ overall ratings of the activity

The Likert scales in the post questionnaire (Figure 6.15) suggest that most participants found the activity useful and worthwhile. Changing the rules was perceived as an interesting feature by almost all participants (24 out of 26 agreed).

| | agree | somew | neither | somew | disagree |
|--|-----------|-----------|-----------|-------|-----------|
| I used our planet's wall display a lot. | 13 | 6 | 4 | 2 | 0 |
| I used the other planet's wall display a lot. | 9 | 6 | 4 | 4 | 2 |
| I used the wall display for the game settings a lot. | 9 | 10 | 4 | 2 | 0 |
| I wish there had been more space to walk around. | 4 | 1 | 10 | 3 | 8 |
| The game was a welcome break in the day. | 15 | 5 | 2 | 4 | 0 |
| The game served as an icebreaker to help people get to know each other. | 15 | 9 | 1 | 1 | 0 |
| The game provoked a useful discussion about the topic. | 6 | 12 | 4 | 3 | 1 |
| I would have preferred a 90 minute lecture on the topic instead of the game. | 1 | 2 | 2 | 9 | 12 |
| The game would be useful before a lecture series. | 8 | 11 | 3 | 4 | 0 |
| The game would be useful after a lecture series. | 6 | 8 | 9 | 3 | 0 |
| Changing the winning condition added an interesting perspective to the game. | 24 | 1 | 1 | 0 | 0 |

Figure 6.15.: Participants’ ratings of the whole activity

Opinions appeared divided regarding whether the activity should be scheduled before or after lectures.

6.14.7. Comments in the debriefing

The debriefing was essentially structured as a Q&A, with participants raising concerns, suggestions and requests for clarification to which the teacher replied. While some of the comments were constructive, others could be interpreted as complaints. For instance,

one participant simply expressed emotional frustration without providing any explicit suggestion for improvement:

- *“On the last round, so our team won by maximising [income]. I feel quite sick from winning that last round. Like we, we raised the temperature to 5 degrees or something [...] and we won! We maximised our money. I don’t know how many billions of people were killed in doing that. I feel quite sick from it.”*

In response to this comment the teacher clarified the activity’s pedagogical intention:

- *“We aren’t saying that that is the way the world works. I’m just trying to give you choices to think about how you might play. How you might think about this balancing act if the world is like this or like that, you know, how it might work out.”*

6.14.8. The teacher’s actions

During rule-changing the teacher generally stayed close to the team that was making the decision. On some occasions he talked to the team or the whole group and sometimes looked at the tablet, as seen in Figure 6.13.

There was one brief instance where the teacher offered assistance by reminding the group what each mitigation scenario meant, while they were trying to make a decision. This intervention was not about the teacher explaining how to use the tablet interface. Rather, the teacher used the structure of the tablet interface as a concrete, auxiliary means to explain some abstract key concepts in the simulation.

The teacher was comfortable managing the new features. Switching between playing and editing the simulation was perceived as fluid and convenient. Apart from the new features, the orchestration load was essentially the same as in the previous session, i.e., easily manageable. The teacher’s only suggestion for improvement was that a clock on the floor would help him keep an eye on the time.

6.14.9. The teacher’s reflections on rule-changing

As the teacher told me excitedly after the session, he found that the questions that were raised tended to be remarkably profound, much more so than in the previous session with the static rules. However, the groups had not always been able to come up with good

answers to those questions. The teacher attributed this to a lack of common background knowledge within the group, which in turn was explained by a lack of prior lectures.

The teacher argued that a 20 minute mini-lecture was not enough time to introduce the topic to a depth that was adequate to the course. For instance, his introduction had barely brushed over the relation to the Stern report and many other important concepts. This explained participants' frequent questions and uncertainty regarding how the mathematics worked.

6.15. Discussion

The above evaluation in a real training course provided empirical evidence of collaborative rule-changing being usable, useful and desirable. The findings are discussed below. First I will reflect on the extent to which the findings support the explicit design objectives. Then I will do the same with the Benefits of Rule-Changing that were proposed earlier to evaluate participants' active and collaborative engagement with the simulation rules. Afterwards I will go into detail about usability, usefulness and desirability.

6.16. Reflections on the design objectives

The following table reflects on the findings with regard to the design objectives that had guided the rule-changing design iteration 6.2.

| N | Design objective | Status | Comments |
|---|--|---|--|
| 1 | Allow participants to apply relevant course knowledge in practice by discussing alternative rules and scenarios | supported with caveats | These discussions engaged the whole team and were consistently on-topic and productive. Informed choices depended on a prior conceptual and procedural understanding of the topic, which some individuals had more of than others. |
| 2 | Ensure that the interfaces take minimal effort to learn and use | supported | - |
| 3 | Ensure that a single teacher is able to facilitate the rule-changing | supported | The discussions were essentially single-stream, allowing the teacher to be aware and get involved when appropriate. |
| 4 | Encourage participants to reflect on assumptions through decision making and discussion | supported in an unexpected way | Yes in the sense that “assumptions” could also refer to newly acquired conceptual understandings. |
| 5 | Leverage existing team structures to facilitate collaboration and give both teams equal opportunities | supported | Rules were changed at the team level. Both teams were allowed to change the rules twice during the session. |
| 6 | Use ambient information to enable whole-group awareness of the rules and reflection on previous rules | supported with potential for design iteration | Potential for future versions to formally support combined winning conditions. |
| 7 | Allow all participants to contribute equitably to the rule-changing and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics | supported with caveats | A lack of prior domain knowledge may have compromised some participants’ ability to contribute equally. Briefly sharing a tablet in a large group (up to 13) worked surprisingly well in terms of physical accessibility. |

Table 6.2.: Relating the findings back to Gameplay Objectives

6.17. Reflections on the Benefits of Rule-Changing

The table below reflects on the findings in relation to the Benefits of Rule-Changing as measures of success (Figure 6.3). All of the benefits were found supported by the rule-changing feature. Two limitations concern number 2 and 6.

| N | Benefit | Status | Comments |
|---|---|------------------------------------|---|
| 1 | Talk about rules as a group, rather than individually | supported | Decisions about the rules were shared in teams. |
| 2 | Discuss, rather than complain | supported with new question raised | During rule-changing the teams discussed, rather than complained. In questionnaire and debriefing comments it was not always possible to clearly distinguish between complaints and constructive suggestions. |
| 3 | Peer-to-peer, rather than student-teacher | supported | The teacher was not called upon during rule-changing, although he stayed nearby to provide clarification if needed. |
| 4 | Focus on alternative assumptions, rather than punishment and reward | supported | None of the comments were framed in terms of punishment, reward or penalty. |
| 5 | Proactive exploration over passive reception | supported | The groups chose different scenarios. New definitions of winning were proposed that exceeded the choices that the application offered. |
| 6 | Engage with the inside of a glass box, rather than the outside of a black box | partly supported | More like opening the lid of the black box a tiny bit, while the walls (i.e. the maths?) were still opaque. Dissecting the rules into scenarios and winning conditions (that could be replaced) provided some high-level insight into the thinking behind the model and game design and should be seen as an important first step. Several participants expressed strong interest to understand the maths in more detail. |
| 7 | Distinguish between model and game | supported | Mitigation/adaptation scenarios and winning conditions were discussed separately. |

Table 6.3.: Overview of the Benefits of Rule-Changing and how they were supported

Regarding point number 2, the findings clearly showed that, given the opportunity (i.e., during rule-changing) the groups preferred action and discussion over complaining. Although the teacher was usually nearby and listening (thus potentially available for receiving complaints), at any time the groups rather engaged with the choices on the tablets and had a focused discussion. This did, however, not imply that discussion became (or should become) the only mode of communication throughout the session. The debriefing was in Q&A form and - like the questionnaires - welcomed constructive feedback as well as complaints. This raises a need for clarification of what benefit number 2 aims to achieve: (a) to eliminate complaints altogether or rather (b) to provide specific opportunities for discussion where discussion is the appropriate mode? In the teacher's and my own opinion, (b) is pedagogically more reasonable. A well-orchestrated activity should provide a time and place for participants' active discussion (e.g., through gameplay and rule-changing) as well as complaints (e.g. in the debriefing).

Point 6 was partly addressed, although the findings indicated much potential for future work. Arguably, an idealistic goal for the application would be make the relevant underlying maths completely transparent and modifiable. Several participants had made suggestions in this direction and a consequent "glass box" approach, as suggested by Jonassen and Strobel (2006) and advocated by Turkle (2003) would call for this. Allowing detailed modifications to the model and game structures might arguably unlock further potential for learning, such as fostering digital literacies by helping participants adopt the language of game designers (Salen 2007; Games and Squire 2008) or providing accessible entry-points into computer programming (Utting et al. 2010). In relation to these advanced goals, this work merely scratches the surface.

6.18. Usability

Using simple tablet menus and ambient information, this work identified a usable way to enable collaborative rule-changing at classroom scale. This is a novel achievement, even though the evidence was limited to a small number of high-level changes and brief (less than 3 minutes) chunks during which participants could endure huddling together around a tablet. Although these short chunks were found appropriate in proportion to the 90-minute session, the question is raised whether the proposed design could scale up to longer periods of rule changing. The latter would be necessary for enabling more detailed modifications to the model and game structures. For instance, future training courses might conceivably include follow-up workshops, dedicating several hours to collaborative

improvements to the climate model. In this case there would be further design questions regarding appropriate representations, interfaces, mechanisms for collaboration and orchestration.

Besides, differences in participants' prior skills might become increasingly relevant the more detailed the modifications become. For instance, at a point where the modifications require advanced mathematical or programming skills it might become more difficult for the design to ensure equitable opportunities for each individual to make informed contributions if the group has diverse skills. This consideration indicates a potential tradeoff between inclusiveness and level of detail.

Arguably, the reasons why the current design was usable involved a variety of factors, including the following.

1. **Individual accessibility**, such as being able to read the text from a metre away in a shared context. This was possible due to large fonts, which in turn were only feasible because the amount of text on each menu page was minimal. This in turn was enabled by having individual menu pages for each question, and letting the teacher (rather than participants themselves) switch from one page to the next.
2. **Easy operations**. The modifications were implemented through simple operations of *replacing* one element (e.g. a scenario or winning condition) with another of the same kind. More involved operations, such as adding, deleting, removing, composing or otherwise manipulating elements, (or perhaps drawing or entering numbers or text), were not needed in this application. Whether such operations might be needed in future designs is debatable. In any case, easy operations were found beneficial for usability, learnability and shareability.
3. **Learnability**. The tablet interfaces were designed to be self-explanatory. Using familiar and descriptive names for the choices and a logical “plain English” screen layout with integrated buttons and numbers, the interfaces allowed participants to reliably infer the function and meaning of each interface element, simply by reading the sentences off the screen. The interfaces were indeed found to be self-explanatory, as participants were immediately able to discuss the provided alternatives by switching between options using the radio buttons. No difficulty was added by the ambient displays, since they used the same representations as the tablet interface.
4. **Shareability**. Up to 13 participants were observed successfully sharing a tablet for a short period of time, up to 2.5 minutes.

5. **Orchestrability** was ensured by (a) avoiding choices that could potentially break the time frame of the session, such as choosing the number of game rounds, (b) letting the teacher (rather than participants themselves) switch from one page to the next, (c) the teacher's ability to intervene when appropriate, by overhearing the singular (!) stream of conversation. The latter was arguably encouraged by the fact that only one tablet was used for rule-changing.
6. **Appropriate integration** into the overall activity. Changing the rules between the matches allowed groups to explore different alternatives rather spontaneously, in response to whatever learning need emerged during gameplay. Since the rule-changing discussions were short, the whole activity stayed within schedule. The ambient displays were usable (and necessary) as reminders of the rules for the current match and previous matches.

6.19. Usefulness

Three aspects were identified that speak for rule-changing as a useful addition to the overall activity.

1. **Additional teacher awareness.** This was an unexpected benefit that resulted from the emergent group dynamic of keeping rule-changing discussions (mostly) to a single stream of dialogue and decision-making.
2. **Whole-team clarification and reflection on understanding.** The rule-changing was found to provide an appropriate context for a whole team to reflect on their understanding together. In contrast to the gameplay situations, which were generally more hectic and often fragmented into multiple lines of talk, here the whole team usually listened when a participant asked for clarification. This arguably increased the participant's chances to get competent answers, due to a bigger circle of listeners. Moreover, the rule-changing challenged the group to predict how a particular choice would affect the optimal game strategy. Getting these predictions right required the groups to apply a reasonably detailed and accurate understanding of the model. These were useful moments for individuals and the whole group to check explicitly whether they all shared the same understanding.
3. **An appropriate context for talking about scenarios and winning conditions** Rule-changing provided a dedicated time and place for discussing scenarios

and winning conditions in a more appropriate context than the contexts offered by gameplay and debriefing. During gameplay, whereby groups think and act “inside” one particular scenario, the discussions and decision-making are relative to that single scenario. By contrast, rule-changing offers a more objective, “outside” perspective - in the sense of Ackermann’s (1996) notion of “diving-in” and “stepping-out”. Arguably, comparing multiple scenarios and winning conditions is easier from this perspective, and helps the groups make informed judgements *about*, rather than *within* the game and the model. This outside perspective is also given in the debriefing. However, an important benefit that rule-changing offers and debriefing does not, is the prospect of being able to try out one of the discussed options immediately. This provides a more empowered context for hypothesising about implications of alternative scenarios, due to the opportunity to test some of the hypotheses through play.

For the above reasons it was decided to include rule-changing as a standard part of the 4Decades activity. While adding unique benefits, rule-changing did not seem to detract from any of the other modes. On the contrary, it can be argued that the rule-changing phase focused on addressing a specific learning need, namely the group’s need to discuss scenarios and winning conditions. By addressing this need in its own dedicated context, the other phases (notably gameplay and debriefing) were “freed up” to focus on their respective strengths. For instance, participants needed the debriefing to relieve emotional stress (Fanning and Gaba 2007) and talk about wider pedagogical implications of the activity, such as whether it should better be used at the end of a course, rather than at the beginning.

6.20. Desirability

Questionnaire and debriefing comments indicated that an overwhelming majority of participants was strongly in favour of rule-changing. This finding corroborated tentative suggestions based on session 1. The teacher agreed that rule-changing offered unique benefits by “*forcing the group to ask themselves these profound questions*”.

Being able to replace the preset scenario and winning conditions with alternative choices allowed the group to express alternative conjectures regarding (a) how they thought the economy works and (b) what climate/economic goals they considered realistic. Making an explicit distinction between these concerns was arguably a further pedagogical benefit

of the rule changing. It separates the “model” from the “game” and allows participants to critique those separately. Indeed, participants’ discussions and comments tended to focus on either one or the other at any time, rather than mixing them up.

To what extent the rule-changing helped participants empathise with the designers’ perspective was difficult to measure empirically under the given circumstances. Presumably, since participants’ exposure to the insides of the simulation was relatively short and high-level, it would be unrealistic to look for specific changes in participants’ use of language - along the lines of Salen’s (2007) study with the Gamestar Mechanic simulation - that would indicate that individuals thought of themselves more as game designers, rather than recipients. Indeed, the choices that participants were offered were superficial compared to the detailed modelling decisions made by the designers. Nevertheless, what the 5 minutes of rule-changing arguably achieved was to allow participants to take ownership of some of the simulation’s key questions. Changing the rules put them in control of these questions - rather than at the receiving end of a (perceived) system of punishment and reward - which doubtlessly is a more empowering and desirable place to find oneself in.

6.21. Conclusions from Analysis 2

Collaborative rule-changing in a classroom-size group was found usable, useful and desirable. As the evaluation in a real training course suggested, rule-changing can be useful for triggering profound questions and to give the groups a sense of ownership of these questions that could not be achieved by merely playing the game and talking about it in the debriefing. Providing participants with a dedicated context for discussing these questions by changing the rules required additional ambient information and another set of tablet interfaces that were easy to learn, use, share, orchestrate and integrate into the overall activity. For the purpose of this design, some of the goals and experience from the previous design were reused and the *Benefits of Rule-Changing* were proposed as a useful tool for evaluation.

The in-situ evaluation showed that groups (of up to 13) were able to make and discuss rule-changing decision together by sharing one tablet and, in the process, benefit from peer clarification and increased teacher awareness. The rule changing was found desirable addition to the overall activity, according to comments from participants, the teacher’s

opinion and qualitative analysis of relevant mechanisms involved in collaborative rule-changing.

The analysis further suggest a number of caveats for future design. Firstly, the evidence suggests that, although a group of 13 could share a tablet for a few minutes, things might get physically uncomfortable with more participants or a longer time span. Secondly, although the simple operations on the tablet (choosing scenarios and conditions using buttons) were sufficient for the given purpose of making high-level changes, supporting more detailed changes might potentially entail more complex operations, which might not be straightforward to design whilst ensuring important use qualities, such as easy usability, learnability, shareability, orchestrability and equitable accessibility to all participants.

In response to the research question guiding this analysis, the above analysis shows that simulation rule-changing *can be feasible* in classroom-size groups. Moreover, it was perceived as very desirable in an adult learning setting when dealing with an advanced topic. The sum of observed benefits warrants strong arguments for the use of rule-changing in future designs.

Future work might address the related question as to whether rule-changing generally increases learning gains. The circumstances of the study did not allow for a controlled comparison of the activity's effectiveness with rule-changing versus without. Such an investigation would arguably require much larger sample sizes.

Chapter 7.

Analysis 3: Continuous and equitable accessibility

The analysis described in this chapter focuses on the difficult challenge of ensuring *continuous accessibility* under real conditions where groups may be moving around dynamically. This issue is particularly relevant to *usability whereby the classroom is the user* (Dillenbourg et al. 2011). Unlike traditional usability studies, here we are dealing with a “user” that has approximately 50 eyes and hands, multiple threads of attention and an amoeba-like ability to move and change physical shape. In pursuing the task goals, this curious creature can even split into parts, take on multiple roles and re-group in different spatial configurations. It can even physically stand in its own way, as the following analysis will show.

7.1. Overview

This chapter reports on all four in-situ sessions. Notable differences between the sessions are highlighted regarding how the simulation was delivered and appropriated by the participant groups. Based on a qualitative comparison of whole-group formations, implications of these differences are discussed and design recommendations are made with regard to accessibility. There are many aspects to accessibility, potentially including physical, cognitive, cultural and other factors. Whereas the importance of prior knowledge has been emphasised in the previous chapter, this chapter focuses on issues of *physical access to peers and technology*, which turned out to be a critical factor in the case of 4Decades.

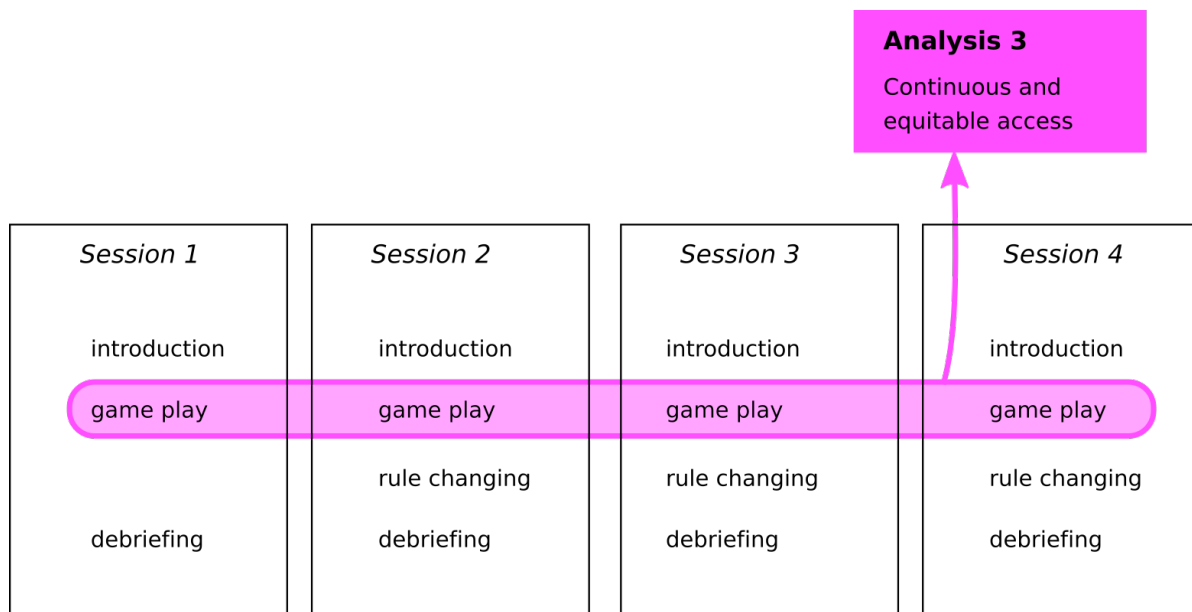


Figure 7.1.: The following analysis looks at *game play* in each of the four sessions.

7.2. Question and aims

The question that guided this analysis was:

Q2-C. How can continuous and equitable access be ensured for all participants?

The following aims were addressed.

- to identify critical issues regarding physical access to peers and technology (from now on simply called “issues” in the context of this analysis) that appeared in the wild with 4Decades; and to explain the causes of these issues;
- to devise useful abstractions and recommendations for design and orchestration to help avoid similar issues in the future.

7.3. Method

The analysis draws mostly on video, audio and photos. Questionnaire data was used additionally on one occasion. The visual material presented in this analysis includes stills from wall mounted video cameras, a handheld video camera and photo cameras. In the figures below, the images are often cropped and annotated in order to help the reader focus on specific aspects of relevance. Sometimes it was unavoidable that other persons

(such as researchers and course faculty) are visible in the background, watching or doing other things. Generally, this should not distract the reader significantly. In fact, in many cases it will be either irrelevant or obvious, depending on the context, that those persons are not participants. However, in the occasional case when explicit clarification seemed appropriate, I tried to provide this information, e.g., in the figure captions.

7.4. Findings

All the relevant issues that were found in all of the four in-situ sessions are pointed out below. After a schematic overview I will describe each session in detail. The only session where severe issues were observed is session 3, as will be detailed below. Session 1 and 2 showed minor issues and session 4 showed no issues.

7.5. Overview

To analyse how the design relates to issues of physical accessibility, it makes sense to look at (a) the designed configuration of technology and furniture in the room, (b) relevant aspects of how participants appropriated this configuration, and (c) issues that emerged as a consequence of (a) and (b). Figure 7.2 summarises (a), (b) and (c) for each of the sessions. The individual diagrams of each session will reappear in the following descriptions and discussions of each session. Here is how to read the diagrams:

- **tables:** large grey rectangles
- **projection screens:** thin grey rectangles
- **chairs as initially provided:** small grey squares
- **tablets:** white rectangles
- **players (i.e. sitting participants):** black squares
- **analysts (i.e. standing participants):** black circles
- **problem areas:** ellipses

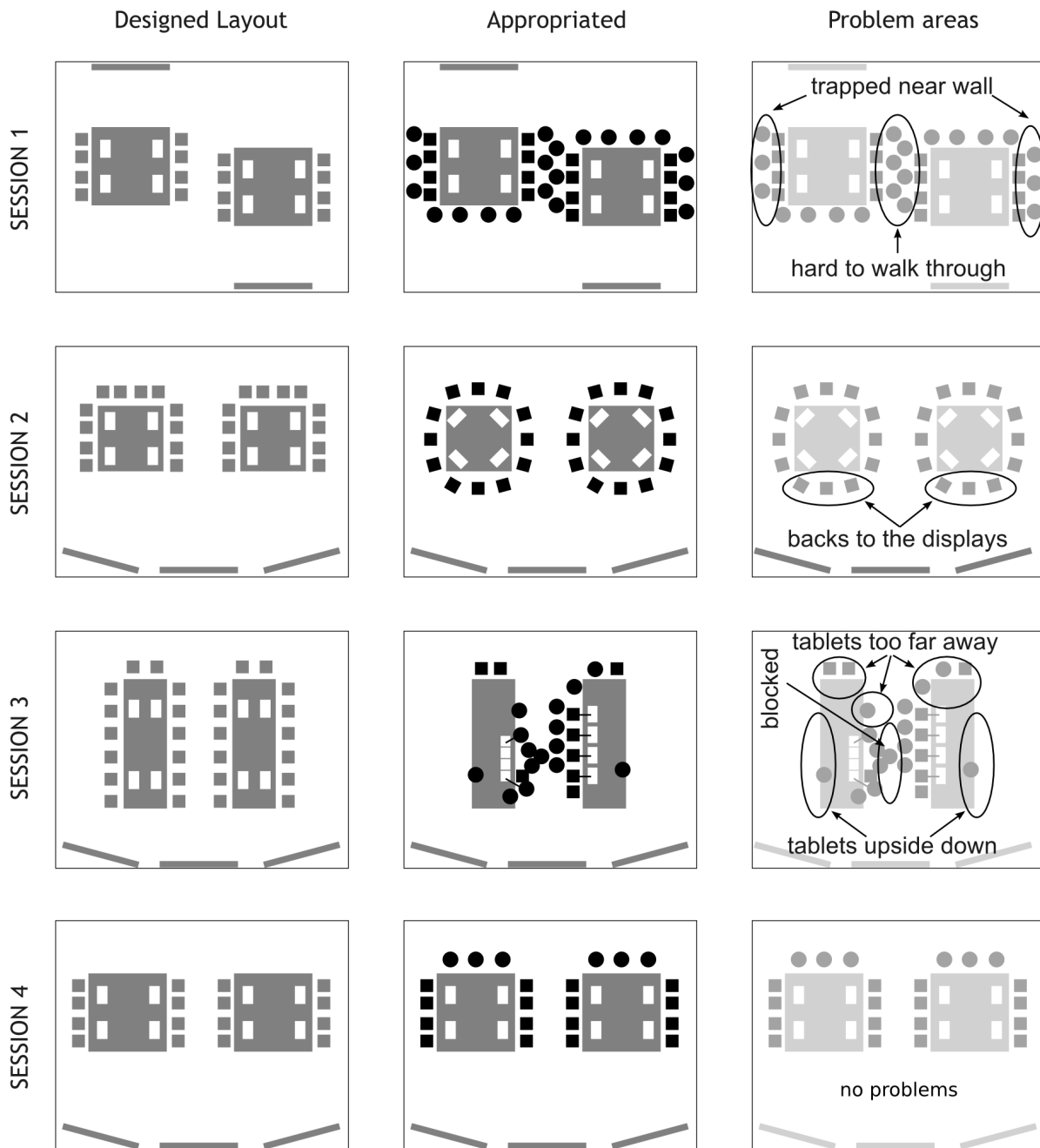


Figure 7.2.: Summary of spatial issues compromising access or comfort. Left column: the layout as prepared at the beginning of the session. Centre: Actual configuration in a typical moment during the activity. Right: Areas where some participants experienced issues with accessibility.

7.6. Session 1

Figure 7.3 shows a schematic overview of session 1 (as seen before in the larger overview in Figure 7.2 above).

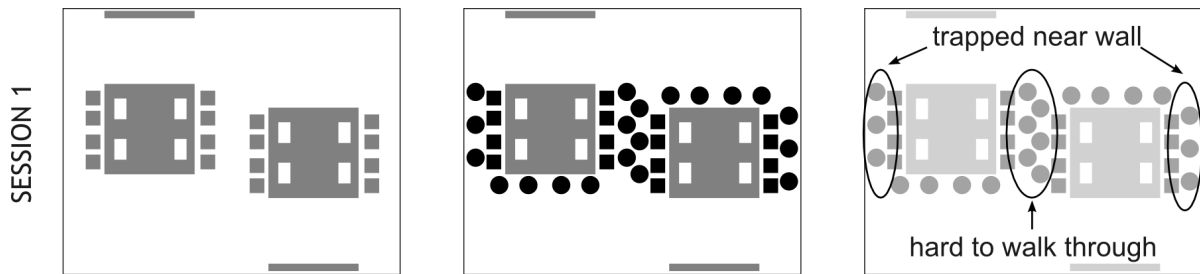


Figure 7.3.: Session 1: Designed layout (left), appropriated layout (middle) and issues (right)

7.6.1. Designed Layout

As described in Chapter 5, the layout in session 1 comprised two square tables, each with four chairs on two sides. The tablets were initially positioned at the edges of the tables, to be used in pairs. The two ambient displays were at opposite sides of the room.

7.6.2. Appropriation

As was the design's intention, the tablets were used in pairs and neither turned nor repositioned. The analysts distributed themselves around the tables, except the space between the tables and the ambient display.

7.6.3. Issues

The room was slightly too narrow for people to walk comfortably around the tables. While this did not affect the players, it occasionally affected the teacher and some of the analysts.

Figure 7.4 shows that the space between the tables was rather narrow, such that the teacher (bottom left) was not always able to walk easily from one side to the other. Moreover, some analysts stood near the walls, somewhat trapped between the wall and the chairs and thus less able to walk around freely. For instance, they could not simply walk over to the other side of the table if they wanted to, e.g., to look at other tablets.



Figure 7.4.: Session 1: Narrow room, people leaning against the walls

Despite this minor impediment, all analysts were able to participate equitably. From any standing position they could easily see one or both ambient displays, and at least one of the tablets, which was more than the sitting participants whose field of view was usually limited to their own tablet and their own team's display. Therefore, both roles (player and analyst) had their advantages.

The above issues could be regarded as minor, since none of the participants were actually excluded and no-one showed or mentioned any physical discomfort. However, the findings indicated that a limit of group size was reached. The group was relatively large in this session, partly because half a dozen faculty also participated occasionally. If there had been many more participants (e.g., more than 30) given the same room size, then conceivably there might have been bigger issues.

The minor issues described above did not occur in Session 4 where the layout was similar but the room was larger and the group was smaller (cf bottom of Figure 7.2).

7.7. Session 2

Session 2 involved three ambient displays next to each other. The additional display was placed in the middle for both teams to see equally well. The team displays were slightly turned inwards. Besides that, the layout was similar to the previous sessions, featuring square tables with tablets at two sides (Figure 7.5).

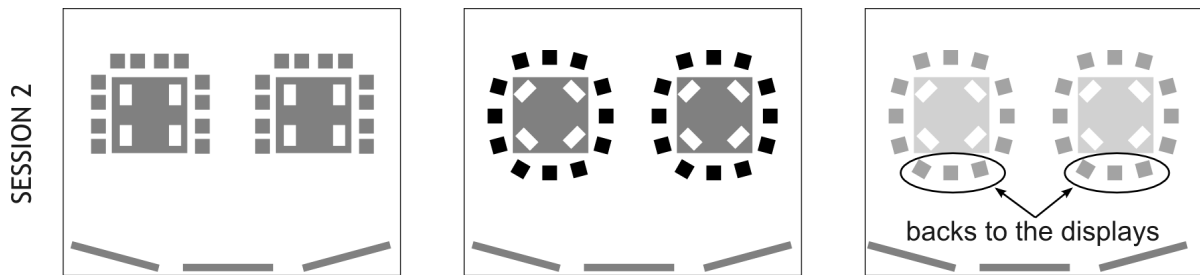


Figure 7.5.: Session 2: Designed layout (left), appropriated layout (middle) and issues (right)

7.7.1. Designed layout

Chairs were provided for all participants this time. Since the group had not met before, the assumption was that they all wanted to be treated equally. Therefore, rather than dividing the group into players and analysts, they were all players.

7.7.2. Appropriation

Counter to the original design, the teams moved their chairs in a full circle around their table. Moreover, they shifted the tablets to the table corners and turned them 45 degrees, (see the middle part of Figure 7.5). This symmetrical formation emphasised equal access to the tablets for everyone, as well as easy talking across the table. Arguably, by moving into a circular formation, the groups made an implicit statement, advocating equal contribution and denying any hierarchies or special roles among the group (Figure 7.6).

Like in the previous session, there was much discussion across the table. Small changes to the seating occurred occasionally, as some participants walked around the table for over-the-shoulder interaction with other regions (e.g., the man in the polo shirt in Figure 7.7). Unlike in the previous session, nobody walked over to the other team or made explicit verbal contact with the other team.

7.7.3. Issues

The only minor issue in this session was that some participants sat with their backs to the ambient displays. This required them to turn around occasionally. The particular participants in this session did not seem to mind turning, as far as the video data showed, and none of the comments explicitly raised the issue. Therefore, with regard to this session, the problem was considered acceptable. Nevertheless, before making a general



Figure 7.6.: Circular team configuration in session 2



Figure 7.7.: Panorama photo of session 2

recommendation for this circular layout, one should keep in mind that participants with diminished physical flexibility might be at a disadvantage if sitting with their backs to the displays.

7.8. Session 3

Session 3 was the only session where serious issues occurred. As the findings show, the proposed layout failed to satisfy the groups' need to collaborate quickly and easily in teams. In response to this, the teams rearranged themselves and the technology in ways that provided some benefits whilst also creating unexpected drawbacks that made it difficult for some individuals to comfortably take part (Figure 7.8).

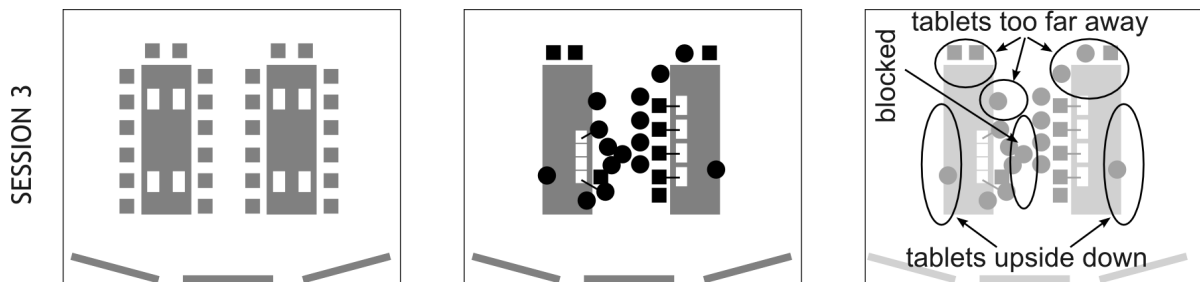


Figure 7.8.: Session 3: Designed layout (left), appropriated layout (middle) and issues (right)

7.8.1. Designed Layout

In this session the tables were much larger. Due to logistical reasons that were not in our hands, the simulation had to take place in the same room as the rest of the course and rearranging the tables was not an option. Therefore we had to work with the given configuration. As usual, the eight iPads were placed near the corner of the tables, facing towards the sides. Like in the previous session, the group was unacquainted, so it was decided to have equal roles (and thus chairs) for everyone. The team sizes were 14 (team A) and 15 (team B).

7.8.2. Both teams: Problematic Corner-L formations

Both teams started off sharing the tablets in small groups of 3 or 4. These small groups initially talked only amongst themselves, focusing on their own tablet, but not

communicating with the other groups. This was consistently observed in both teams and also in the other sessions.

What stood out in this session was the different ways in which the small groups configured themselves and their tablets at the table. Rather than all participants moving closely together side-by-side, as observed in session 2, the regions stayed separate from each other, each region at its own corner of the table. Four general types of formation were identified among the total 8 regions:

1. Corner-L
2. Symmetrical corner
3. Four in a row
4. Three in a row

All four types are exemplified in Figure 7.9. In the left image, what seems like an equitable group of three participants is actually a group of four in the formation that I call “Corner-L”. The fourth person, not shown on the left, can be seen in the larger view on the right image: the woman with the pink sweater at the very top. Her head is turned because she is reading the tablet from a 90 degree angle. In this case, the rotation of the tablet favours the majority (here, three other participants), whereas a minority (here, one person) is at an angle like the short end of the letter L. The person at the short end is at a clear disadvantage with regard to using the tablet. By contrast, the tablet in the upper left corner (same image) is rotated such as to achieve a compromise for all members of the small group. This formation, which I call “symmetrical corner”, is reminiscent of session 2 where equal access was emphasised and symmetrical corners were the prevalent formation across the whole group. Actually, this particular example could also be interpreted as a Corner-L, single one side appears more dominant and the tablet’s angle is not really 45 degrees. “Four in a row” can be seen at the bottom and “three in a row” at the right hand side of the image. Both formations were observed as conducive to equitable sharing.

Team A showed similar formations (Figure 7.10). Again the regions initially work in isolation. The three women on the left shared their tablet equitably as “three in a row”. At the bottom of the left image, four men are in an Corner-L formation, 3:1. Like in the earlier example, one participant is “angled out”, requiring him to tilt his head in order to read more easily. This participant did not touch the tablet and was much quieter than



Figure 7.9.: Regions in team B working in different formations. No talk or eye contact between the regions in the beginning

the other three. Moreover, he did not join as the other three started discussing across the table (right image in the figure).



Figure 7.10.: Regions in team A starting off in isolation (left) and later talking across the table (right)

Figure 7.11 shows two more corner formations observed in the same team. The left one was symmetrical with three participants. The one on the right was a Corner-L, with the participant at the top being outnumbered and out-angled by the three on the right.

Remarkably, both teams in this session ended up migrating to a different formation that was more whole-team oriented, rather than isolated regions. However, the resulting formations were notably different and had implications for the groups' experience. Below I will first describe and explain the transformation on team A, then do the same with team B, before comparing and summarising the issues on both teams.



Figure 7.11.: Symmetrical corner (left) and Corner-L (right)

7.8.3. Transformation in team A

The transformation in team A happened 26 minutes into the simulation, just as the teacher was about to open the second match. While the rest of the team was still quietly listening to the teacher's announcements, one participant initiated the transformation by persuading the other regions to move the tablets together in a row. This act of persuasion is represented in the following excerpt of dialogue. The initiating participant is referred to as "A" and highlighted in bold.

- **A:** *"Should we not get the iPads in one row to get everybody to see [...]"*
- One of the addressed participants starts passing the tablet over towards A.
- A turns to address another region, explaining the suggestion while gesturing with both hands, outlining a wide rectangle on the table.
- **A:** *"Hey, before we get organised, shall we just - because we're one planet - shall we have the four iPads and four drivers sitting side by side."*
- B: *"Yeah. Because we [unintelligible. Passes the tablet to A]"*
- **A:** *"Who's gonna be the driver? We need four drivers and you need to be sitting next to each other."*
- C: *"Yeah that's fine with me."*
- **A:** *"Then come here"*
- A's neighbour: *"Come here"* [stands up to offer her seat]

- Papers handouts are moved out of the way. Tablets are passed across the table. Participants walk around the table to one side.

As the team moved around, four participants - one from each original region - sat down at the tablets as drivers. Seven stood behind them (from now on called the navigators). One navigator stayed in her chair next to the drivers and three participants stayed behind at their original positions - presumably partly because there was little space in the central area. In this new formation each driver had one tablet in front of them and was usually the only person to touch their tablet (see Figure 7.12 below).



Figure 7.12.: Four drivers in a row. Adjacent drivers showed much face-to-face awareness of each other but communication did not span the whole row end-to-end.

7.8.4. Implications for team A

Adjacent drivers and navigators engaged with each other frequently by talk, eye contact, pointing and looking at adjacent tablets, as the above figure furthermore illustrates. This kind of exchange was particularly intense around the middle of the row of tablets, as the navigators argued about team-wide decisions. Drivers occasionally engaged in these arguments, sometimes unsure whose advice to follow. Although adjacent drivers showed continuous awareness of each other and each other's tablet screens, there was barely any direct communication between drivers that sat further apart.

Another issue that was caused by the linear tablet layout can be seen in Figure 7.13. As people turned 90 degrees to look at the ambient display, they ended up standing behind each other, rather than next to each other. Consequently, few participants had an unobstructed view of the ambient data.

Finally, it is worth observing that all the hotspots were condensed in one narrow, overcrowded area in the centre. Once all the limited "hotspots" were taken, some people were stuck with the remaining "coldspots" that offered poor access to peers and technology. Figure 7.14 shows two men in the foreground for whom the tablets were at a difficult



Figure 7.13.: Lucky if you're tall: Many people struggling to access the ambient information at the front of the room

angle and distance. Since every available spot meant being excluded, could the woman sitting in the back be blamed for not really engaging?



Figure 7.14.: “Coldspots”: everywhere but in the overcrowded centre

7.8.5. Transformation in team B

Team B undertook a similar transformation, also initiated by a participant. It happened at the start of the third match (minute 39, one match after team A's transformation).

One participant suggested, while standing up and pointing, *“Can we just come over to this side”*. As the tablets got passed around, another participant commented (shrugging and in a tone that might have passed as very mild sarcasm) *“So that we can all sit together? Well, I don’t mind. Just get all in one place, so we can all sit - and do we know which [tablet] is which then? Does it matter?”* Without explicitly objecting, this comment subtly hinted at three potential issues: Firstly, was it physically realistic for all 15 team members to gather on one side? Secondly, might the team potentially lose overview of the planet data? Thirdly, had the team already made an implicit agreement about handling all the regions equally? Indeed, all three concerns turned out to be justified, as indicated by the findings described below.

The resulting configuration was a **driverless row of tablets** (Figure 7.15). Although the arrangements of tablets in a row was superficially similar to the other team’s configuration, the critical difference here was that this team had not nominated any drivers, which raised the unexpected challenge of how to negotiate control. This challenge appeared to overwhelm the team as a whole. The only person sitting at one of the tablets had been sitting there before and did not take a leading role. Generally, the whole team appeared unsure what to do.



Figure 7.15.: A driverless row of tablets. Left: The team is undecided. Right: Still undecided. Several hands in pockets.

7.8.6. Implications for team B

To explain this undecidedness, several factors may potentially have played a role, including (a) a lack of roles to structure the interaction and (b) an overwhelming amount of information in one place. Unfortunately the data did not provide a definitive answer in this regard. However, some participation issues were apparent:

1. One participant experienced difficulty reading the tablets upside down across the table.
2. Three disengaged participants at the top end of the table (not counting the man in the suit behind them, who is faculty.)
3. One man behind the main group was completely trapped between the teams, unable to see the tablets over the shoulders of the standing group. (His head can be spotted by following the black arrow at the top right in both images.)

7.8.7. Impact on participant experience

Arguably, team B suffered more severe consequences of their team's reconfiguration than team A. Whereas part of team A was intensely engrossed - particularly the drivers and the handful of navigators right behind them - team B appeared rather dysfunctional as a whole and having less fun. This suggested the impression that team A, despite some apparent disengagement and exclusion, had a more satisfying experience as a whole than team B. This suggestion was supported by the video data and also by team A's consistently higher average ratings on all the Likert scales in the questionnaires (Figure 7.16).

The figure shows abbreviations of the Likert scale statements. The full wordings are listed below.

1. I enjoyed the game.
2. I used our planet's display a lot.
3. I used the other planet's display a lot.
4. I used the display for the game settings a lot.
5. I wish there had been more space to walk around.
6. I thought the game was a welcome break in the day.
7. The game served as an ice-breaker to help people get to know each other.
8. The game provoked a useful discussion about the topic.
9. There was enough time for discussion between the matches.
10. I wish the calculations behind the simulation had been explained in more detail

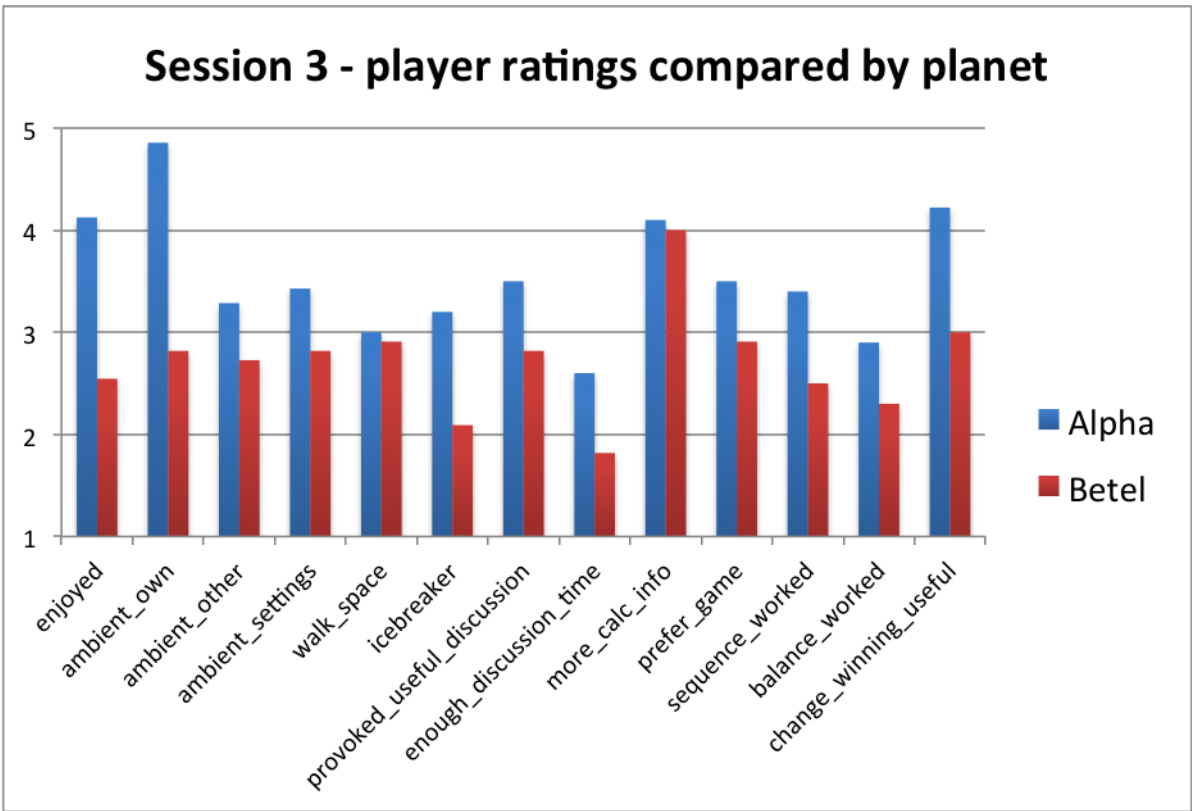


Figure 7.16.: Average ratings were consistently higher in team A (Alpha Centauri) than team B (Betelgeuse) on Likert scales where 5=agree, 4=somewhat agree, 3=neither agree nor disagree, 2=somewhat disagree, 1=disagree.

11. I would have found it more useful for learning about the topic if we had skipped the game and instead spent the two hours with just lectures and discussions. [inverted scale]
12. The sequence of lecture, game and discussion worked well for me.
13. The balance between lecture, game and discussion worked well for me.
14. Changing the winning condition was useful for learning about the topic.

7.8.8. Summary of session 3

The session started with a Panopticake layout of tablets, except that chairs were provided for all participants, just like in session 2. However, whereas the groups in session 2 had appropriated the design by creating a perfect circle of “symmetric corner” formations, this group took a different turn. Initially they worked in small groups near the corners. Then, frustrated by the large distance between the regions, they moved all the tablets to one side; thereby essentially defeating the Panopticake’s intention to allow everyone within a team to see what everyone else is up to (i.e. mutual awareness). The motivation for this formation was explicitly stated in team A as allowing everyone to see the tablets simultaneously. This plan evidently was not completely successful, as the narrow space and large table effectively prevented everyone from seeing the tablets reasonably comfortably. This problem was the same in both teams and resulted in some participants being excluded.

Remarkably, the 5 participants that noticeably suffered from L-Corners turned out to be the same 5 participants who noticeably disengaged later on in the game. No such disengagement was found in any of the other sessions, which were free from L-Corners.

7.9. Session 4

Figure 7.17 below shows a schematic overview of session 4.

7.9.1. Designed Layout

The layout in session 4 was similar to session 1 in that the tables were square and 16 chairs were provided for players, leaving the rest of the group to start off as analysts.

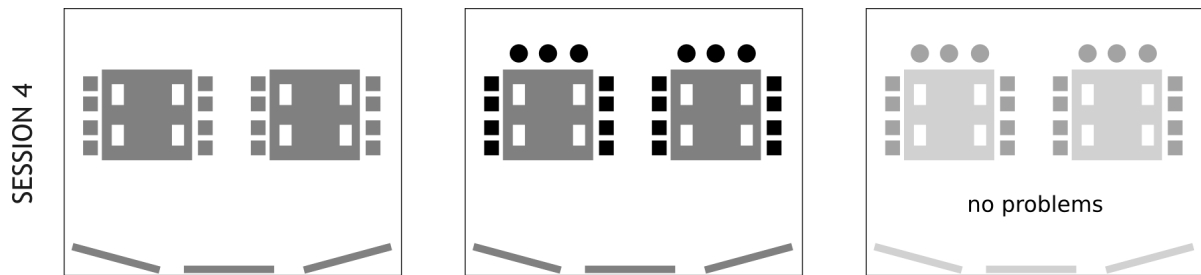


Figure 7.17.: Session 4: Designed layout (left), appropriated layout (middle) and issues (right)

7.9.2. Appropriation

Since the group was rather small in this session (23) and the 16 chairs were constantly used, there were only 3-4 analysts at any time. The room was large and plenty of space was available for the analysts to move freely around the tables. An overview is shown in Figure 7.18.



Figure 7.18.: Session 4: Two square tables, three projectors and a large clock on the floor for the teacher (bottom left)

In this session there were no observed issues of accessibility whatsoever. The analysts had plenty of space to walk around and mostly stood at the rear sides of the table where they blocked no-one's view. As seen in Figure 7.19, the Panoptcake layout was appropriated just as intended by the design, with pairs sharing the tablets equitably and teams facing each other. Like in session 1 the tablets were neither moved nor turned.

7.9.3. High overall satisfaction

The whole session worked without any issues and produced remarkably high results in participant satisfaction. 15 out of 19 questionnaire responses agreed that they enjoyed



Figure 7.19.: Session 4: The simulation was used as intended.

the game, while the other 4 ticked “somewhat agreed”. Moreover, the majority stated that they found the game useful for reflection and discussion and they found rule-changing crucial for their learning (Figure 7.20). Moreover, the teacher found that this session was the most successful and satisfying of all four sessions.

| | | agree | somew | neither | somew | disagree |
|--|--|-------|-------|---------|-------|----------|
| I enjoyed the game. | | 15 | 4 | 0 | 0 | 0 |
| The game helped me reflect on what I learned in lectures. | | 11 | 8 | 0 | 0 | 0 |
| The game provoked a useful discussion about the topic. | | 11 | 7 | 1 | 0 | 0 |
| There was enough time for discussion during the session. | | 6 | 6 | 4 | 3 | 0 |
| Changing the game settings was crucial for my learning. | | 14 | 2 | 3 | 0 | 0 |
| I could have learned the same from a board game. | | 1 | 1 | 2 | 8 | 7 |
| I could have learned the same from a PC game. | | 3 | 3 | 6 | 4 | 3 |
| A different game would have been more useful for learning. | | 0 | 0 | 10 | 6 | 3 |

Figure 7.20.: High overall satisfaction with 4Decades in session 4.

7.10. Full overview of appropriation

An high-level overview of all four sessions is presented in the following two diagrams. Both diagrams resulted from the above analyses of how the design was appropriated by participants and the teacher in the wild. Figure 7.21 summarises relevant concepts and relations that were found in all sessions. Figure 7.22 focuses on the specific issues in session 3 that explain the collaboration breakdown in both teams.

The vertical axis is divided into 4 parts that describe instructional mechanisms, learning mechanisms, peer social interaction mechanisms and technological mechanisms. The horizontal axis represents a spectrum of group size: individual, pairs or small groups, teams, whole-class. Starting at the bottom of Figure 7.21, some examples are provided below, demonstrating how the figure should be read:

- Ambient information was used individually, in pairs or small groups, in teams and as a whole classroom.
- The tablet interfaces were used in pairs and small groups.
- The arrows indicate a logical direction of purpose. For instance, social comparison (right, yellow), was used at the whole-class level to support reflection.
- Chains of arrows can be read like this: The tablet interfaces supported negotiation by sharing control over the interface.
- The ambient information was used to support several of the teacher's actions, including ad-hoc lecturing, linking to previous lectures, debriefing, and game facilitation.
- Ambient information was used directly to support reflection.
- Ambient information was used by the game teacher to encourage social comparison
- The tablets were used for making decisions, which in turn changed the tablet contents in real time.

Figure 7.22 highlights two notable differences in session 3. Both differences concerned the use of tablet interfaces.

1. The tablets were **used by individuals** in some cases, contrary to the design's original intention.
2. Using the tablets **as a whole team** was attempted but not successful (dotted lines).

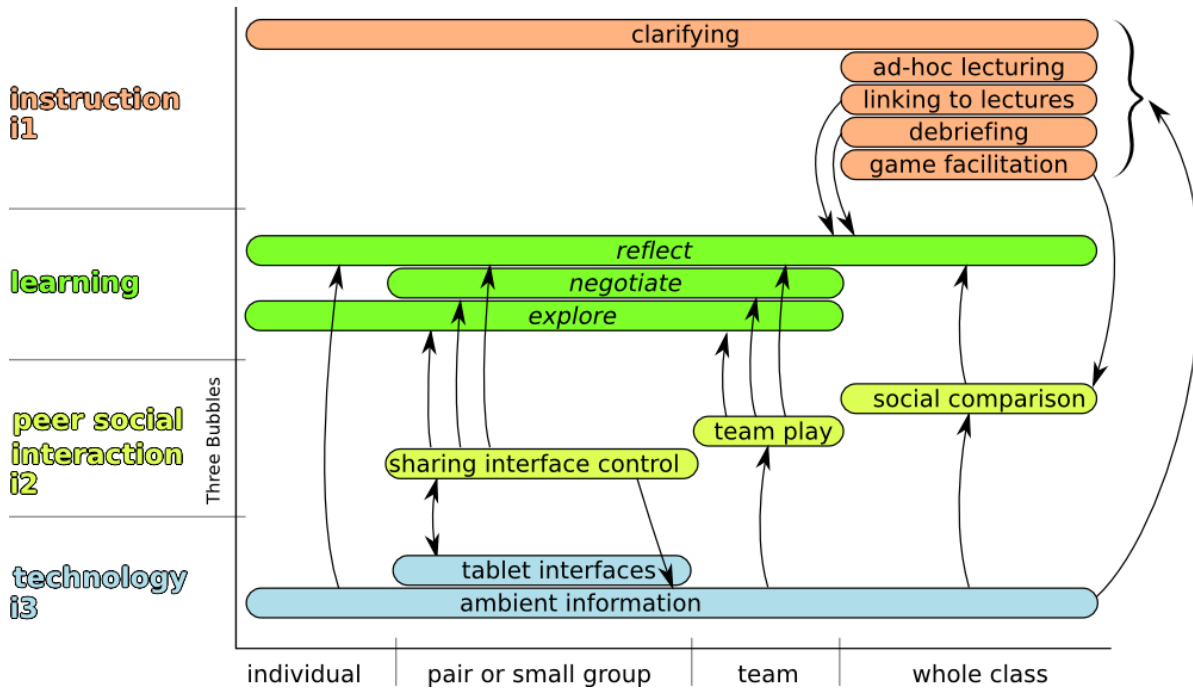


Figure 7.21.: Appropriation in the wild: Commonly observed relations in all sessions

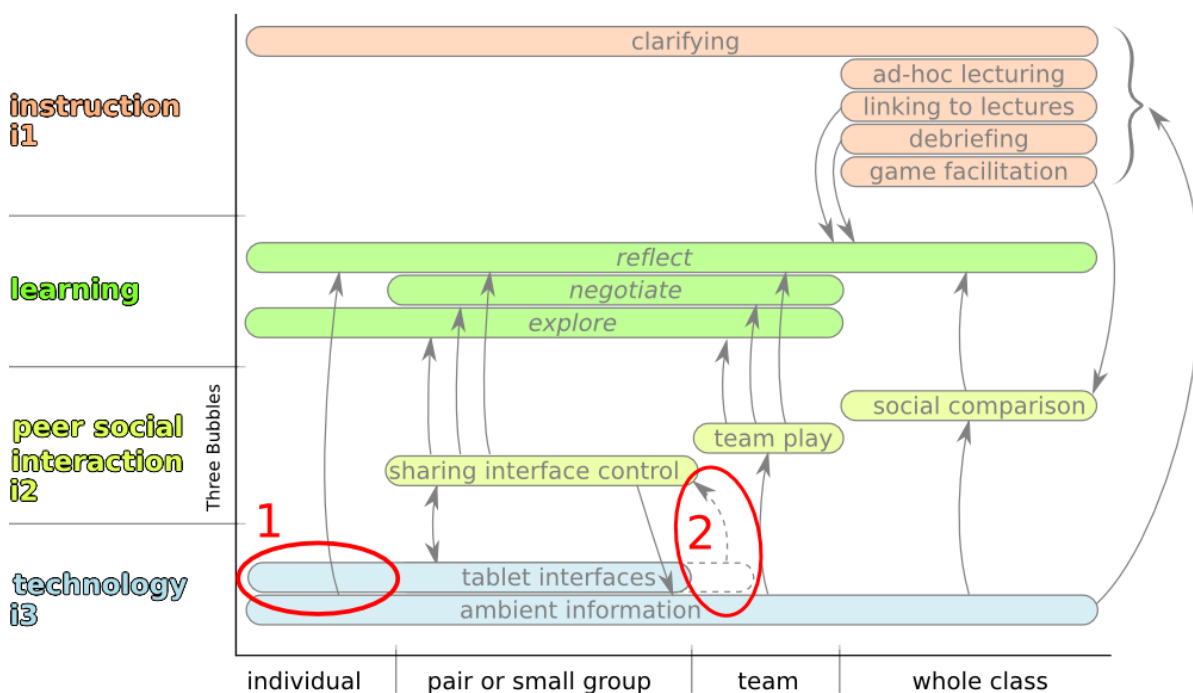


Figure 7.22.: Appropriation in the wild: Issues specific to session 3

7.11. Discussion

Four sessions were analysed with regard to issues of physical access to peers and technology. Detailed video analysis identified a variety of relevant issues in this context and highlighted the importance for design and orchestration to consider emergent aspects of large-group dynamics.

7.12. Effects of group formations on mechanisms of collaboration

In the following, the mechanisms for collaboration (Yuill and Rogers 2012) framework is used to evaluate how well the observed group formations supported collaboration. Small group (region) formations and large group (team) formations were compared separately. The former are evaluated in regard to *control over the interface*, whereas the latter are judged by all three aspects of the framework: *control over the interface* (here: tablets), *availability of background information* (here: being able to see the ambient displays), and *mutual awareness of actions and intentions*.

7.12.1. Implications of small-group formations

Looking at how participants shared the tablets in pairs or small groups (regions), the following small-group formations were observed in situ. Figure 7.1 summarises them in relation to how well they supported equitable control of the tablet interface.

7.12.2. Implications of large-group formations

Regarding how large groups (teams) configured themselves and the technology in a spatial context, the following large-group formations were observed in situ, as summarised in Figure 7.2. The figure also includes each formation's typical implications as observed in situ.

| Formation | Control | Observed in session(s) |
|--------------------|--|--------------------------------|
| Single driver | individual | 1 occasionally, 3 in team A |
| Pair | equitable | 1 and 4 consistently |
| Three in a row | equitable | 3 initially |
| Four in a row | equitable | 3 initially |
| Symmetrical corner | equitable | 2 consistently, 3 occasionally |
| Corner-L | consistently observed in association with physical exclusion | 3 initially |

Table 7.1.: Overview of observed small-group formations and their implications

| Formation | Control | Information | Awareness | Observed in session(s) |
|---------------------------|--|-------------------------------|---|------------------------|
| Panoptcake | equitable | high for navigators | high | 1 and 4 consistently |
| Circle seating | equitable | high | high | 2 consistently |
| Long table seating | encourages problematic Corner-L formations | high | high within regions, low across table | 3 initially |
| Row of drivers | individual | unequal due to standing crowd | high between neighbours, low across distance | 3 in team A |
| Driverless row of tablets | unequal, confusing | unequal due to standing crowd | unequal due to standing crowd (visual and physical obstruction) | 3 in team B |

Table 7.2.: Overview of observed large-group formations and their implications

7.13. Recommendations for design and orchestration

The above analysis highlighted the importance of physical access to peers and technology for successful collaboration in classroom-scale. In this context, a range of conducive and detrimental spatial formations were identified, based on detailed video analysis. Recommendations are made in favour of formations that were consistently observed to support

essential mechanisms for collaboration (namely *control over the interface*, *availability of background information*, and *mutual awareness of actions and intentions*). Conversely, formations that tended to compromise these mechanisms are not recommended.

For sharing tablets, recommended formations are: (a) **Pair**, (b) **Three in a row**, (c) **Four in a row**, and (d) **Symmetrical corner**. No recommendation is made for **Single driver** because, although this formation did not lead to any obvious issues per se, it failed to exhibit the benefits that the other sharing methods showed (such as triggering immediate conversation as soon as the tablets become interactive). Finally, it is recommended that future designs should definitely avoid the **Corner-L** formation, since this formation was shown to cause spatial exclusion of individuals with remarkable consistency.

With regard to large groups (teams), two formations can be recommended: (a) **Panoptcake** and (b) **Circle seating**, since both were found to support equitable control and awareness. A minor difference between the two is that **Panoptcake** involves standing as well as sitting participants, whereas in **Circle seating** all participants are continuously seated and have equal roles, i.e., everyone is treated equally. Due to the latter feature, **Circle seating** might potentially be preferable with groups where social factors are particularly sensitive for various reasons, such as personal, interpersonal, cultural, political, developmental or professional circumstances. For instance, some unacquainted groups of adult learners can be challenging in this respect, as the teacher emphasised based on personal experience. By contrast, **Panoptcake** lets participants take turns between sitting and standing, thus offering more variation and a potentially more physically refreshing experience. This may be desirable in some cases, for instance, with acquainted groups that are tired after many hours of lectures. Depending on the group size and room layout, sitting participants in this formation may sometimes be less able to see the other team's ambient display, since the view might occasionally be blocked by standing participants. However, this disadvantage is counterbalanced since (a) sitting participants have the advantage of using the tablets and (b) participants swap roles after each match.

Problematic large-group (team) formations that should be avoided, based on our findings, are **Long table seating**, **Row of drivers**, **Driverless row of tablets**. All three were shown to have inherent detrimental impacts on some of the mechanisms for collaboration. Symptoms of these detrimental impacts included (a) overcrowded areas, (b) excessively large "coldspots", (c) long distances between regions that are supposed to collaborate face-to-face, (d) Corner-L formation, (e) groups or individuals looking somewhat lost,

undecided or frustrated. Recognising any of these symptoms in situ should prompt teachers or designers to check whether there might be issues of group formation.

Tables should be small enough, allowing teams to talk easily face-to-face in all directions across the table. Enough space should be ensured around the tables so that standing participants can move around freely without crowding. If this cannot be provided then perhaps a seated formation (Circle seating) might be preferable. Projectors should be either wide angle or ceiling mounted to ensure that the displays are sufficiently large, well-positioned and well-oriented to ensure equal access. The tablet screens should be of reasonably high quality, allowing participants to read them from a wide range of angles.

7.14. Explaining severe issues that caused exclusion

The only *severe* issues were observed in session 3 where the format of the tables offered poor support for whole-team awareness and coordination. In an attempt to overcome this problem, the teams invented alternative formations which turned out to cause other issues. Although the resulting large-group formations (Row of drivers and driverless row of tablets) worked well for some participants, other participants ended up being excluded.

As far as the video analysis suggested, these participants did not choose to be excluded, nor did others consciously exclude them. Rather, it appears that continued exclusion was mostly a function of starting off in an unfortunate position. This alarming finding is supported by the observation that in session 3, all 5 participants who started off at the disadvantaged side of a Corner-L (and only these 5) showed clear signs of frustration and disengagement - and did so throughout most of the session. A plausible explanation is that initial frustration discouraged these participants from putting effort into the (arguably futile) struggle of merging with the crowd. Other participants were partially excluded for other reasons, such as being stuck in the middle of a crowd, practically unable to get to their team.

A conclusive explanation was found in emergent group dynamics that were nurtured by physical factors of the room, particularly the size and format of the tables and the positions of the technology. Moreover, participants seemed largely unaware of these dynamics and their implications. This leads us to another, summarising recommendation:

Generally, designers and teachers working with shared technology should never assume that a group of more than two participants will automatically organise itself into a

spatial formation that suits its needs. What the group needs and what the emergent spatial dynamics encourage may be different things. As our findings have shown, groups' self-organisation can sometimes result in genuine improvements (such as the Circle seating that emerged in session 2) but it can also go horribly wrong (see session 3), leading to some individuals being excluded from taking part.

7.15. Learning about inclusion by studying exclusion

Being excluded was naturally unfortunate for the affected participants, since their enjoyment and potential learning experience were compromised. It is worth emphasising that excluding participants had not been the intention of this research; nor could the teacher or participants be held responsible, because the observed incidents of exclusion were not easy to predict. Nevertheless, these incidents were eventually of value to the analysis, by helping us better understand (and predict) the success or failure of large-group collaboration with shared technology. To better understand what keeps the classroom-scale activity intact, it can be helpful to examine when, where, how and why it broke apart.

The first thing that broke in session 3 was equitable control of the tablet interface (due to Corner-L formations, affecting 5 participants). The fact that the consequences were undesirable in all 5 cases supports the suggestion that equitable *control over the interface* was not just beneficial but crucial (in line with Yuill and Rogers 2012).

The second thing that broke in session 3 was *mutual awareness of actions and intentions*, as one end of the **Row of drivers** formation in team A did not know what the other end was doing. The breakdown in team B - a collective state of paralysis surrounding a **Driverless row of tablets** - could be seen as a combination of both those mechanisms, since the team ended up unable to negotiate (a) *control over the interface* and also lacked (b) *mutual awareness of actions and intentions*, the latter in this instance being a consequence of lacking apparent actions and intentions per se.

The third thing that broke in session 3 was *availability of background information*, as many participants were unable to see the ambient display, due to a dense crowd of participants standing in front of each other.

The observations described above corroborate the usefulness of Yuill & Rogers' (2012) framework for explaining the successes and failures of collaborative tasks, by looking at

three key mechanisms *control over the interface*, *availability of background information*, and *mutual awareness of actions and intentions*. Two relevant claims that this analysis contributes in this context are that (a) these mechanisms can also be applied in large-group settings and (b) large-group spatial dynamics can in turn be critical to the success or failure of each of these mechanisms. What follows is that there is a need for useful conceptual tools that make large-group spatial dynamics easier and more efficient to diagnose and correct.

Indeed, this analysis has identified a range of issues in this category that are potentially threatening to the goals of inclusive design. Some of these issues are inherently subtle and hard to spot. Particularly, the issue of Corner-L formations and its severe implications only came to my attention after many hours of intense video analysis. That a teacher in the heat of classroom action is likely to miss these subtle signs among a large group must not surprise anyone. To help teachers and designers sensitise their perception to these important issues, I made an attempt to describe indicative symptoms to watch out for during an activity. The above list of recommended and discouraged formations (small and large) is complete with regard to the 4Decades data but may be extended through future design and evaluation.

The ultimate responsibility to ensure conducive spatial conditions in a session (by spatially organising tables, chairs and tools) lies with the teacher. This has been argued by (Dillenbourg and Jermann 2010) and is corroborated by our observation that (a) designers lack control over how a group might change a given layout and (b) a group's well-intended choices in this regard may not always be well-informed and hence not always beneficial. Teachers, on the other hand, can only dedicate a small fraction of their limited realtime resources to managing spatial issues. The orchestration load in this particular respect can be broken down into three challenges:

1. **Knowing configurations that (don't) work.** Arguably, the above list provides a reasonable starting point.
2. **Being able to diagnose spatial issues quickly in real time.** The Three Bubbles Model (described below) aims to help in this regard, by offering a small set of abstractions that is simple enough for teachers to use in real time.
3. **Intervening appropriately in case of an issue.** In what way a teacher might choose to address an issue once it is detected, may vary between circumstances. It may not be reasonable to give generic advice in this regard, certainly not within the scope of this research.

The following section presents the Three Bubbles model which, among other things, addresses point (2) of the above list.

7.16. The Three Bubbles model

To provide a useful tool for evaluation in this context, I propose the Three Bubbles model as a result from the above analysis. Its key premise is that large-group, face-to-face collaboration is best supported by leveraging different *mechanisms for collaboration* at different scales within the social and physical space. The proposed scales are *class*, *team* and *region*, which are instantiated as social and physical contexts for interaction (bubbles) that are nested hierarchically. I.e., the *class* is divided into *teams*. Each *team* is divided into *regions*.

In the terminology of the Three Bubbles model, *region* refers to the smallest unit of digitally-mediated collaboration in the whole activity. It comprises (a) one interface that affords input to the simulation and (b) two or more participants that share control over the interface. Thus a region delineates an entity within the social and physical space in the classroom where a small group of participants has control. This aspect relates to the Latin root *regio* (derived from *regere*, to rule) which is part of the rationale for choosing the word “region” in this context. It is a deliberate choice, rather than simply copying the word from 4Decades, where it originally described geographical regions on a simulated planet, i.e., a topic-specific concept. As part of the Three Bubbles model, which aims to be generic across topics, a region could represent anything within a simulation that has agency, from an ant to a person, organisation or country.

Unaffected by individuals’ ability to move opportunistically and join and leave different regions, it is assumed that regions and teams can be practically treated as cohesive units (socially and physically). At least, if a teacher points and says “this region” or ask a question to a team, it should generally be clear what group is being addressed. This assumption is warranted by the analysis suggesting that effective large-group formations are characterised by devices staying in more or less fixed positions.

Figure 7.23 highlights specific functional relations between the three mechanisms for collaboration (left) and the three types of bubbles (right) that were found most relevant for explaining successful whole-class collaboration. The figure reads as follows:

1. Background information becomes available as simulation data is interpreted.

2. Mutual awareness of actions and intentions is supported by (a) comparing simulation data, (b) negotiating team strategies and (c) sharing devices.
3. The interface is controlled by operating the devices.

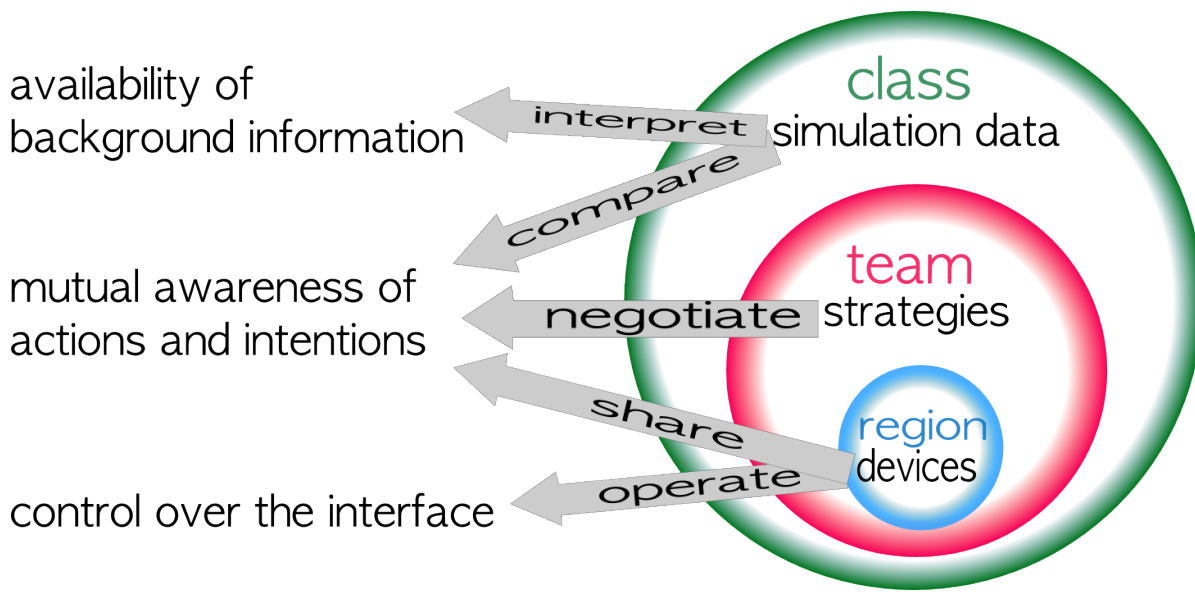


Figure 7.23.: Three Bubbles: mapping mechanisms to different scales of social / physical space

The last sentence (3) above may appear somewhat trivial at first sight, but in combination with (2) it highlights the fact that the interfaces can potentially be operated individually, although this may come at a cost of mutual awareness. Another important aspect are the ownership relations:

1. **Simulation data** is public to the whole class (via ambient displays).
2. **Strategies** are made in teams.
3. **Devices** are controlled by the region (as opposed to, e.g., individuals or the whole team).

From the combination of (2) and (3) emerges a deliberate ambiguity about which bubble is ultimately in charge of decisions. On the one hand, having control over their device allows regions to explore and act independently of the team. On the other hand, team-wide overall narratives and winning conditions encourage teams to act as one. Whenever there was a conflict of intentions between several regions within a team in 4Decades, verbal persuasion was the only viable means for resolution. This rationale of encouraging face-to-face discussion by constraining control was consistently supported by findings from all 4Decades sessions. The Three Bubbles model explains why this “trick” worked,

based on generic mechanisms that are known to support collaboration. Based on this generic explanation, the model predicts that the trick will work across different topics in future designs that successfully implement the Three Bubbles hierarchy as described.

The model represents a high-level abstraction of the key concepts and relations that notably contributed to successful collaboration in 4Decades - successful meaning inclusive, engaging and equitable. Its aim is to foster similar use qualities in other classroom-scale activities, by (a) suggesting well-informed improvements to existing applications, sketches and prototypes; and (b) inspiring ideas for new activities.

7.16.1. Relation to proxemics

The Three Bubbles concept can be distinguished from the concept of *proxemics* (Hall 1962). Proxemics investigates how humans tend to interact at various levels of interpersonal distance, based on observed cultural norms. The idea of proxemic 'zones' (intimate, personal, social and public) has been used in HCI in the context of people and multiple devices (Greenberg et al. 2011). It is relevant to Three Bubbles insofar as distances between people and devices can inform participants' expectations regarding how to communicate and use the system appropriately. However, whilst proxemics aims to explain all kinds of life situations, Three Bubbles focuses specifically on classroom settings with shareable devices, i.e., a relatively narrow range of group sizes and social, physical, and technological constraints. Moreover, Three Bubbles aims to shape, rather than merely explain, certain aspects of group interaction, such as discussion and equitable participation.

7.16.2. Use in evaluation

Analysing face-to-face interaction in large groups can be a tremendous challenge, due to sheer group size, complexity and ambiguities in communication, including verbal utterances, gestures, posture and gaze.

When evaluating a particular design, the Three Bubbles model aims to help its users watch out for certain things happening at certain scale. For instance, if social comparison was observed to happen only within regions, then that would indicate that something is wrong with the design. Rather than providing evaluators with a long list of all the things that might go wrong, a concise checklist of *success indicators* is probably more

useful. Based on the analysis, the most concise answer I can give to the question what a session generally looks like when things go well, is the following rule of thumb: Regions should share, teams should be aware, and the class should compare.

In addition to checking the mapping between behaviours and scales, the model also supports testing the intended hierarchy, i.e., whether the bubbles are properly nested. In particular, this implies verifying that all devices are *shared* (rather than used individually), furthermore that each *region* operates as part of a *team* and that each team engages in social comparison. In the schematic example below (Figure 7.24) the left diagram illustrates a “healthy” activity according to the model, whereas the diagram on the right exemplifies a subtle problem: one sharing group, although engaging in social comparison, fails to collaborate with the rest of its team. Problems like these can be very hard to spot in dense video data, unless one knows precisely what to look for. For illustration, looking back on some of the figures in Analysis 3, the reader may find some aspects of the Three Bubble models reflected in the ellipses drawn around groups in action. The concepts of class, team and region provide a concise and generic vocabulary for describing group activity at nested levels.

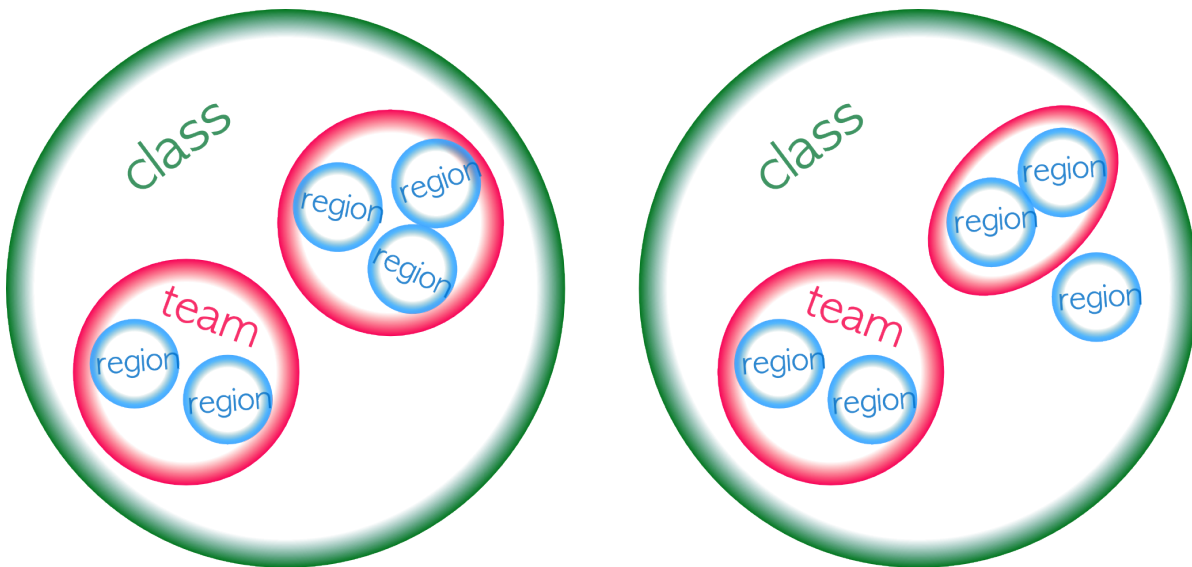


Figure 7.24.: An example: nested bubbles as expected (left) and a breakdown in teamwork (right)

7.16.3. Use as a design template

To the extent that iterative design methods are used that involve continuous evaluation, a tool that supports evaluation, by extension, also supports design. Nevertheless, it

makes sense to ask whether Three Bubbles is also sufficiently prescriptive for the purpose of devising new activities from scratch.

Assuming that any topic worth discussing could potentially be translated into a classroom-scale simulation - and given a specific topic - would Three Bubbles provide a designer with all the instructions necessary to perform this translation with confidence? Arguably not. Three Bubbles represents a useful abstraction with regard to *one* important aspect of the big picture, namely how to integrate collaboration at multiple simultaneous scales of group size. More specifically, it focuses on the specific context of gameplay - as opposed to, e.g., rule-changing, for which there may be more appropriate specific abstractions.

Unanswered questions include how to design the interfaces, how to manage time and how to include the teacher. Although 4Decades can potentially serve as a worked example in these regards, demonstrating how these challenges were resolved in a particular case, appropriate abstractions can provide additional value. More abstractions - regarding other relevant aspects - are needed to support new designs. To address this need, the CLOVER framework (Chapter 8) aims to provide a more comprehensive suite of conceptual tools, of which the Three Bubbles model is a part.

7.16.4. Use as a metaphor for shared user experience

The “bubble” metaphor has been used in other connotations, negatively hinting at the way technology can sometimes isolate people inside their individual digital bubbles. For instance, who has not seen friends at a restaurant ignoring each other whilst engrossed in their private mobile apps? By contrast, here I propose a different interpretation of “bubbles” as immersive contexts that can be shared face-to-face.

Three Bubbles outlines three such contexts, each representing a different type of shared focus, intentionality, or simply “being together”. By making these contexts nested and complementary to each other, a well-designed activity potentially allows groups and individuals to change focus between the contexts fluidly and opportunistically, enabling an inclusive and engaging overall activity.

To summarise, the Three Bubbles model aims to help debug classroom-scale activities by providing useful abstractions for test specific indicators of successful collaboration. Although arguably not comprehensive enough to function as a standalone framework for new designs, the model’s apparently easy usability and usefulness for visual analysis suggests that it can add value to evaluation and iterative design.

7.17. Summary of the third analysis

This analysis investigated what design and orchestration can do to ensure continuous accessibility in a large-group setting with technology. Zooming into details of small-group and large-group spatial formation, the analysis showed how these can change dynamically and affect individuals' and groups' experience in positively or negative ways.

Recommendations were made for design and orchestration, raising awareness to the importance of continued, inclusive, physical access to peers and technology. Notable breakdowns in this continuity were analysed and related to Yuill and Rogers (2012). Potential for future work was outlined to help teachers and designers identify and remedy emergent large-group issues effectively in the spirit of inclusive design. The proposed Three Bubbles model represents the first step in providing a concise set of practical tools for evaluating activities at the scale of 4Decades.

7.18. Overall discussion of the three analyses

The three analyses resulted in the following answers to the research questions:

1. Orchestrating 4Decades in real training courses is feasible overall. Among a range of beneficial design factors, three were highlighted as particularly crucial to the success of the activity: linearity, performance-independent pace and decoupled team play.
2. Teams of up to 15 participants can collaborate on making informed changes to the simulation rules by sharing a tablet interface. Seven *benefits of collaborative rule-changing* were identified as a useful tool for evaluation in this context. Although the devised interface was feasible for simple rule-changing mechanisms, future work should investigate the design of appropriate interfaces for more advanced rule-changing and modelling tasks in collaborative, large-group settings.
3. To ensure continuous and equitable access for all participants, orchestration must pay attention to spatial configurations of peers and technology. The Three Bubbles model provides useful abstractions to explain successes and failures in 4Decades sessions and to guide future evaluation, with regard to physical and social space. With regard to other relevant aspects, such as time and teacher involvement, further abstractions would be useful to inform the design of similar activities.

In the following I advocate the design of “multi-plane groupwork”, whereby collaborating groups are connected via peer social mechanisms, rather than isolated.

7.18.0.1. A case for multi-plane groupwork

The worked example of 4Decades demonstrated feasible methods to support peer collaboration at multiple scales of groupwork (or “planes” in Dillenbourg’s orchestration terminology), from pairs up to the whole-classroom level. Related work has emphasised the importance of being able to *switch* between different planes, e.g. between groupwork and whole-class interaction (Higgins et al. 2012; Kharrufa et al. 2013). Switching is also used in 4Decades. However, in addition to switching, 4Decades introduces a novel and fluid way of *distributing* peer social interaction across multiple planes simultaneously. Particularly, while multiple regions share an interface together, their work is integrated in realtime with a larger unit of collaboration (i.e., the team), and similarly, multiple teams are integrated into a classroom-wide context of social comparison. Since this particular kind of groupwork involves everyone at multiple planes simultaneously, the label “multi-plane groupwork” seems appropriate (MPG from now on).

The Three Bubbles model explains how the mechanisms of *device sharing*, *team awareness* and *social comparison* together provide the interactional “glue” between those planes that enables MPG. This results in vertical bars when expressing a 4Decades session in traditional CSCL notation (Figure 7.25).

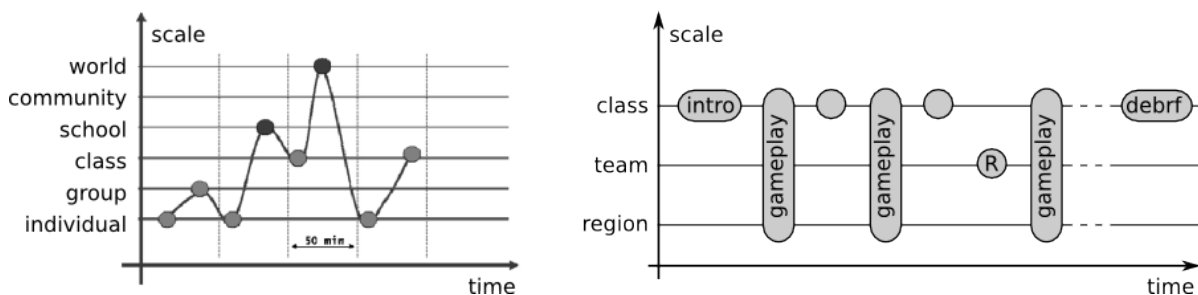


Figure 7.25.: Left: Example of typical CSCL script notation (adapted from Dillenbourg & Jermann 2010). Right: 4Decades gameplay notated as MPG. “Intro” stands for introduction, “R” for rule-changing and “debrf” for debriefing.

A crucial difference between MPG and “normal” groupwork is that the latter typically involves groups working in isolation, whereas MPG essentially connects everyone with everyone else via the above mechanisms. As a consequence, there was practically not a single spot near any of the devices and displays during 4Decades gameplay that was not

observed by peers for more than a few seconds. In addition to explicit communication (e.g., talk across the table) there was also evidence for high peripheral awareness. For instance, the “meander” vignette (see Chapter 5) illustrates how a small group’s private little adventure was detected by members of the larger team. Although each small group was essentially free to make their own decisions, the small groups that belonged to the same team were intrinsically incentivised to look out for each other, at least occasionally and peripherally. Low-level verification that each group was contributing meaningfully thus came from peers, allowing the teacher to focus on higher-level issues. By contrast, in the multi-tabletop groupwork scenario described by Kharrufa et al. (2013), a small group secretly did unrelated, private things whenever the teacher was not looking their way. This example recapitulates a well-known problem in classroom teaching (especially groupwork), namely how ineffective a single teacher’s divided visual attention is as a sole provider of whole-class low-level monitoring. Peer social mechanisms in well-designed MPG, on the other hand, can provide this basic monitoring very effectively and effortlessly, as the case of 4Decades demonstrates.

This raises questions in what cases, and to what degree, whole-class mutual awareness is a desirable and feasible goal to support. In the case of 4Decades it had been an explicit design goal - informed by a concrete learning need within an existing course curriculum. This goal was achieved by introducing subtle incentives for small groups to look out for each other. These incentives derived from a number of implicit constraints in the design, including shared team goals (i.e., winning) and shared resources in the simulation (temperature). Future designs may identify other useful constraints that can be used alternatively or additionally. It is worth noting, however, that adequate room layouts (such as the Panoptিকে) - although doubtlessly conducive - were *per se not sufficient* for encouraging awareness between small groups. Indeed, regions were consistently observed working in isolation before realising that they could perform better in the game by coordinating their actions on a higher plane.

To summarise, the notion of multi-plane groupwork (MPG) emphasises that groups in classrooms can work *independently* without being *isolated*. Arguable benefits of MPG include a tremendous reduction of the teacher’s low-level monitoring load, allowing him or her to focus on high-level monitoring. The case of 4Decades demonstrates that this is feasible and the Three Bubbles model explains how it can be achieved.

7.19. Summary of Chapters 5-7

This chapter reported on how the 4Decades simulation was evaluated in the wild. The activity allowed participants to apply and reflect on lecture-based knowledge through decision-making and debate. The activity's overall feasibility and high participant satisfaction provided empirical support for the pedagogic approach of conducting simulation and discussion after lectures (see Chapter 2). Moreover, the pedagogical method of encouraging conflict resolution through argumentation was supported in a large-group context, by giving regions control over their interfaces, whilst encouraging decision making in larger teams.

The empirical evidence corroborates prior findings in the context of active learning, including the benefit of prior lectures to establish low-level conceptual knowledge (Silvia 2012) and the importance of debriefing for joint reflection, emotional relief and making comparisons to real-life events (Fanning and Gaba 2007). In addition, the newly designed rule-changing task allowed groups to actively explore a variety of relevant what-if scenarios, such as, what if oil became very expensive in the future - how would we expect the planet to then react to particular policies? Making these decisions together forced the groups to ask themselves some profound questions about the topic and consider alternatives. Tablets and ambient information were used to make this task feasible and integrate it effectively with the overall activity.

Based on extensive video analysis, recommendations for design and orchestration were made, specifically addressing emergent large-group issues of using the physical space in relation to technology. Analysing what caused certain physical / social group formations to succeed and others to fail led to the introduction of the Three Bubbles model. This model describes how large-group collaboration can be structured effectively, using relevant mechanisms for collaboration that are leveraged at multiple scales of group size. The model proposes a concise set of abstractions that are independent of the original simulation's topic and thus aim to be generally applicable in future evaluation and design.

Reflecting on common problems with classroom groupwork, it is suggested that the Three Bubbles model could in some cases inform viable solutions to reduce the teacher's orchestration load, by using peer social mechanisms for providing low-level clarification and monitoring. However, questions were raised about whether the proposed model could support the design of activities that are usable and sustainable in more constrained settings, such as school classrooms.

The next chapters will address these questions. Particularly, Chapter 8 proposes the CLOVER suite of conceptual tools to guide future design and evaluation and Chapter 9 describes how some of these tools were used to create sustainable designs in the context of school settings that were more constrained.

Chapter 8.

The CLOVER framework

8.1. Overview

This chapter presents the Collaborative Learning Orchestration for Verbal Engagement and Reflection (CLOVER) framework. The framework comprises a heterogeneous suite of abstractions (conceptual tools and recommendations) that were derived by generalising from the 4Decades findings. By presenting these abstractions in an accessible format, CLOVER aims to guide future design, orchestration and evaluation of decision-based simulations for whole-classroom use. The next chapter presents two design studies to validate the applicability of CLOVER in different design contexts (Figure 8.1).

8.1.1. Research questions

Q3. What abstractions can we identify that would form a helpful conceptual framework for design and orchestration?

- a) How can we describe in abstract terms the essential factors that make or break a discussion-oriented, whole-class simulation?
- b) What practical tools and recommendations can be given to teachers and designers?

8.1.2. Research aims

The aim of generalisation is to create a *useful* framework that is *usable* in concrete design situations. To what extent CLOVER satisfies these criteria will be tested in Chapter 9.

| | chapter 4 | chapters 5-7 | chapter 8 | chapter 9 |
|------------------|--|--|--|--|
| Question | <p>Q1</p> <p>How can the whole group engage with the topic?</p> <p>a) What types of engagement goals are desirable and feasible?</p> <p>b) What mechanisms can effectively support these goals?</p> | <p>Q2</p> <p>How should one orchestrate the activity?</p> <p>a) What constitutes the orchestration load?</p> <p>b) How should the orchestration load be balanced between teacher and technology to achieve feasibility?</p> | <p>Q3</p> <p>What are useful abstractions to aid future design, orchestration and evaluation?</p> <p>a) How can we generically describe the factors that make or break the design?</p> <p>b) What are practical recommendations for teachers and designers?</p> | <p>Q4</p> <p>Are these abstractions also useful when designing for sustainability?</p> <p>a) Is the CLOVER framework applicable in different settings?</p> <p>b) What design factors are critical for achieving sustainability in the wild?</p> |
| Study | 4Decades (lab) | 4Decades in the wild | | UniPad Comfy Birds |
| Data | lab video, field notes system logs teacher interviews student questionnaires debriefing transcripts | in-situ video, field notes system logs, screenshots teacher interviews student questionnaires debriefing transcripts | findings from 4Decades and literature | notes from design meetings with teachers teacher interviews field notes system logs |
| Analysis methods | lab-based prototyping qualitative analysis | detailed video analysis qualitative analysis of system screenshots | reflection on design development of a conceptual framework | lab based prototyping preliminary feasibility evaluations in the wild qualitative analysis |

Figure 8.1.: This chapter in the context of other chapters.

8.1.3. Chapter structure

This chapter focuses on showing the framework, i.e., bringing it together in one place, rather than presenting any new data or discussion. In the following I will explain the rationale and format of the framework and clarify the difference between tools and recommendations. The rest of the chapter represents the tools as graphical figures (including a full-page summary) and the recommendations as tables that are grouped into categories.

8.1.4. Why call it a framework?

In HCI a framework is defined as “*a set of interrelated concepts and/or a set of specific questions that is intended to inform a particular domain area [by helping] designers constrain and scope the user experience for which they are designing. [Frameworks] can come in a variety of forms, including steps, questions, concepts, challenges, principles, tactics and dimensions.*” (Rogers 2012).

By emphasising their direct utility for concrete design, Rogers distinguishes frameworks from higher-level concepts, such as theories (i.e., explanations of phenomena), theoretical approaches and perspectives and paradigms (i.e., adopted research practice). Similarly, the CLOVER framework aims to provide a set of abstractions that make it easier to generalise from successful cases, whilst staying on the practical side of things, rather than getting too philosophical.

8.2. The CLOVER Suite

CLOVER comprises *conceptual tools* and *recommendations* (Figure 8.2). The 8 conceptual tools (also referred to as *tools* for brevity) can be seen as smaller sub-frameworks that bundle a group of interrelated concepts. In brief, a tool proposes a certain way of thinking about a substantial part of the design space, whereas a *recommendation* typically pertains to an individual decision or aspect to which it aims to raise awareness or give advice. Their brevity, concreteness and heterogeneity makes them similar to *atomic patterns* (Prieto et al. 2013), a format that emphasises being easy for teachers to use.

The following sections present the tools and recommendations. The descriptions of the tools are deliberately brief, partly to keep the presentation concise and partly because all

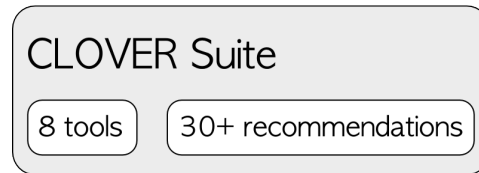


Figure 8.2.: CLOVER comprises 8 tools and many recommendations

of the tools were already described in detail in previous chapters (see cross-references). The figures and tables are mostly the same, except for some occasional minor adaptations that make the framework more consistent. For instance, the terms “small group” and “large group” in Spatial Formations (as introduced in Chapter 7) were renamed here to “region” and “team”. This arguably makes the tool conceptually easier to connect with Three Bubbles. The Gameplay Objectives were extended with some abbreviations to make them faster to read.

8.2.1. Tools

Table 8.1 summarises the 8 tools. The next page presents a graphical overview that is suitable for printing and putting up on a wall. This might be useful as a shared reminder during collaborative design.

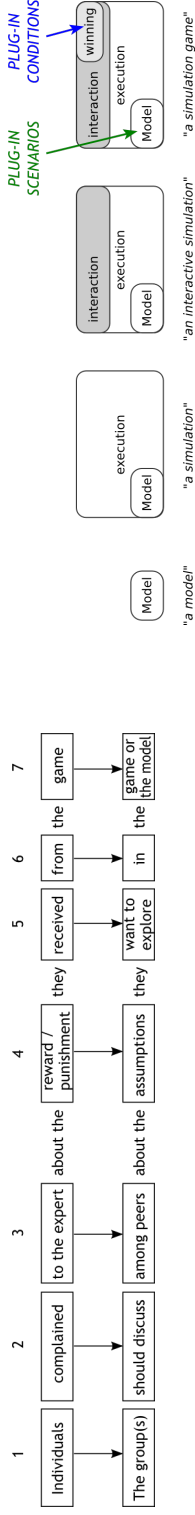
| N | Name | Rationale | Type |
|---|---------------------------|---|---------------------------------------|
| 1 | Model Sim Game | A simulation is the execution of a model. A simulation game is an interactive simulation with a winning condition. Scenarios can be plugged into models. Winning conditions can be plugged into a game. | Activity structure / Definition |
| 2 | Lecture Sim Discussion | Cycles of lecture, simulation and discussion should start with lectures. | Pedagogical strategy |
| 3 | Triple-i Hierarchy | To address a concrete learning need, design should first look at instructional factors, then group interaction and technology last. | Design strategy |
| 4 | Gameplay Objectives | Allow participants to apply relevant course knowledge in practice through decision-making and debate; Design an intuitive multi-user interface for 20-30 participants, etc. | Design goals / Heuristics |
| 5 | Three Bubbles | Divide the class into teams and regions. Distribute information and control to support mechanisms for collaboration across multiple scales | Activity structure |
| 6 | Multi-Plane Groupwork | Encourage awareness and enable appropriately constrained interaction between groups. | Activity structure |
| 7 | Spatial Formations | Teachers are responsible for organising the spatial relations between tables, chairs, groups and technology. Some configurations tend to be more conducive than others. | Orchestration guidelines |
| 8 | Rule-Changing Benefits | An empowered and collaborative approach is preferable when grappling with simulation rules and questions of realism. | Heuristics |

Table 8.1.: The CLOVER tools

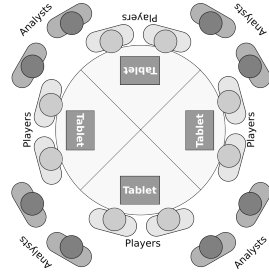
CLOVER

Tools

Classroom Orchestration for Verbal Engagement and Reflection



Benefits of Rule-Changing

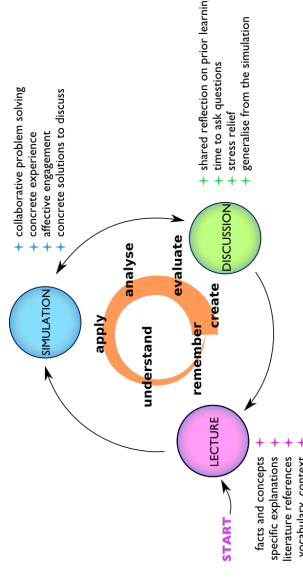
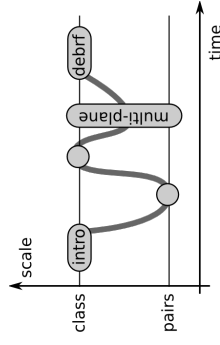


Model Sim Game + Plugins

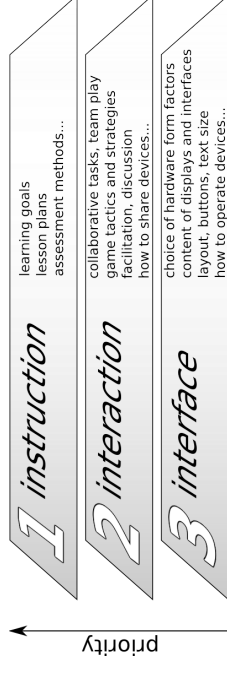
Lecture Sim Discussion

Spatial Formations

Multi-Plane Groupwork

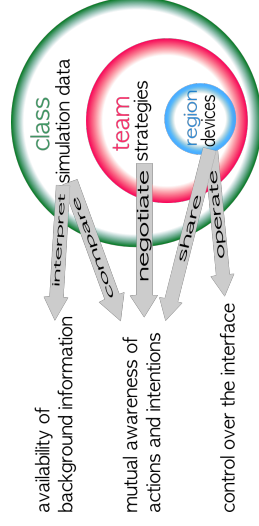


Triple-i Hierarchy



Gameplay Objectives

- 1 apply knowledge
- 2 instantly usable interfaces for 20-30
- 3 single teacher
- 4 reflect on assumptions via decisions
- 5 refresh by moving and talking
- 6 team incentives for negotiation
- 7 social mechanisms and play
- 8 ambient visualisations for awareness
- 9 inclusion and equitable accessibility



8.2.1.1. Model Sim Game

This tool provides compatible definitions for “model”, “simulation” and “game” (Figure 8.3). According to the definition in Chapter 1, Section 1.6.7, these concepts can be briefly summarised as follows:

- a model is a usable representation of the essential aspects of a system;
- a simulation is the execution of a model;
- a (simulation-based) game is an interactive simulation that has a winning condition.

“Pluggable” scenarios and winning conditions were proposed in Chapter 6, Section 6.4.1. Whereas scenarios pertain to a model, winning conditions are specific to games.

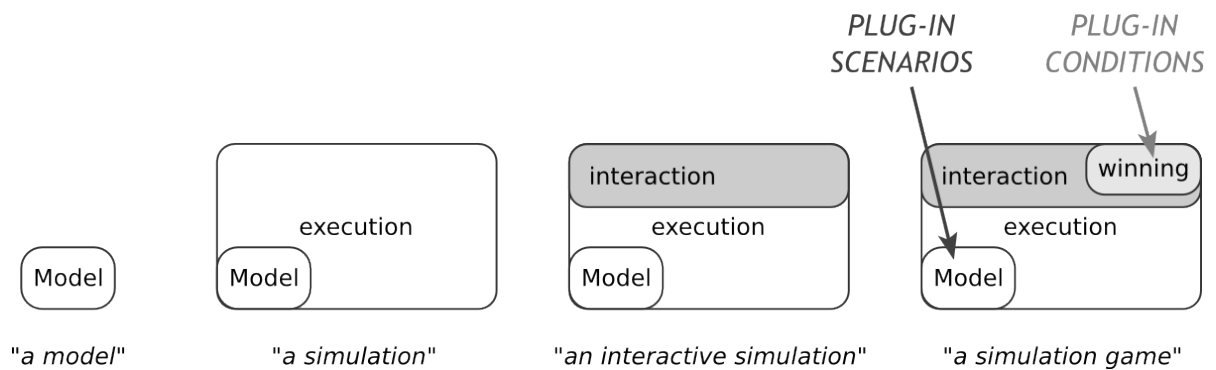


Figure 8.3.: Distinction between models, simulations and games.

8.2.1.2. Lecture Sim Discussion

Lecture Sim Discussion (Figure 8.4) was proposed in Chapter 2, Section 2.3.3. The essential recommendation is for educators to provide lectures prior to conducting a simulation. The purpose of lectures should be to establish an appropriate amount of topic-specific, conceptual knowledge that allows participants to make sense of the model during the simulation. Debriefing should be conducted - in an appropriate form and amount - after every simulation cycle.

8.2.1.3. Triple-i Hierarchy

The Triple-i Hierarchy (Figure 8.5) aims to guide an iterative design processes that prioritises *instructional* goals (i1). Mechanisms of human-human *interaction*, such as collaboration and team play (i2) are considered subsidiary to instruction. Finally, *interface*

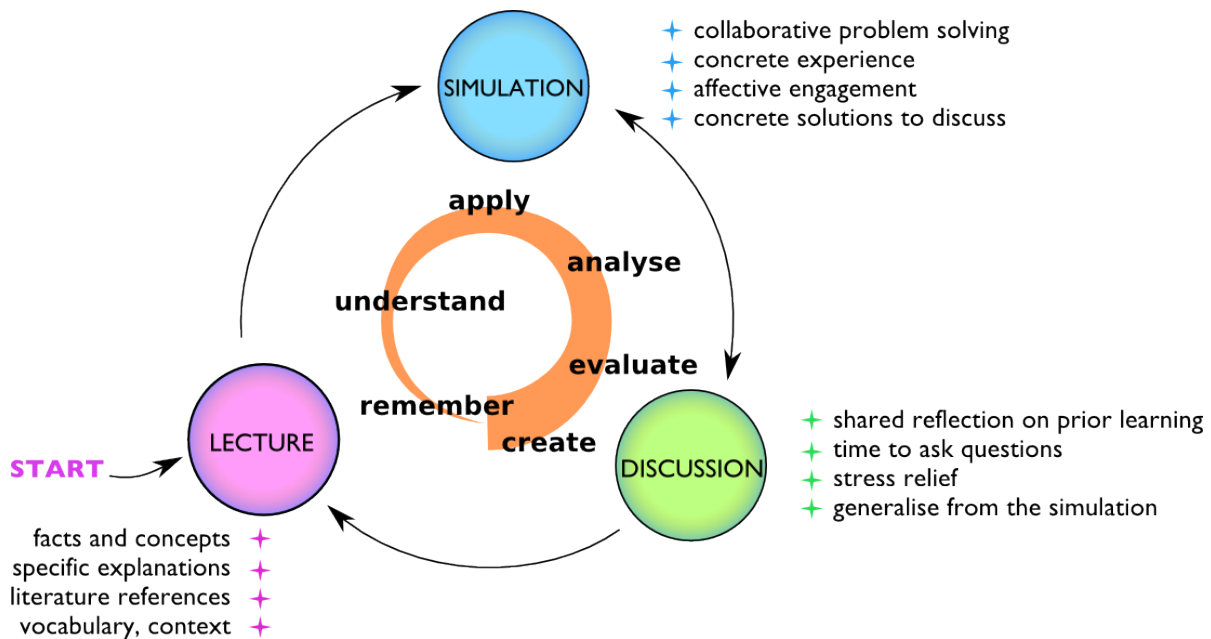


Figure 8.4.: Repeated cycles of lecture, simulation and discussion are proposed to support a continuum of low-level and high-level learning objectives.

design (i3) should support these interactions. The rationale behind Triple-i is detailed in Chapter 3, Section 3.1.2. Its main claim is that interaction does not derive from an interface - rather, *interfaces* support *interactions* that are conducive to *instruction*.

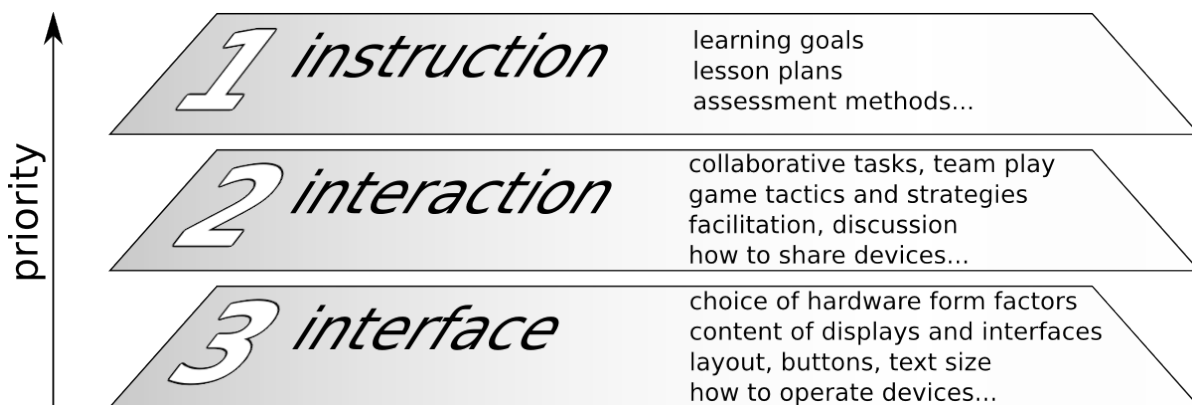


Figure 8.5.: The “Triple-i” hierarchy of design goals

8.2.1.4. Gameplay Objectives

The Gameplay Objectives (Table 8.2) comprise a list of 9 testable goals that were formulated in Chapter 4, Section 4.4.1. Their original purpose was to guide the lab-based design of 4Decades and to test the application's overall feasibility. Moreover, the objectives aim to be useful in a broad range of design cases, including different topics, target groups, settings and learning goals (within the overall theme of verbal engagement and discussion). For this purpose the objectives were formulated in an abstract way, outlining some relevant overall use qualities, but without making detailed prescriptions regarding how to achieve them in a specific case. They aim to benefit iterative design in two ways: (a) to help designers set relevant and achievable goals and (b) to help evaluate - through large-group user testing - the extent to which the intended goals were achieved as the design was appropriated by users in a real context. Finally, the Gameplay Objectives aim to be *compatible but independent*, thus allowing designers to use some of the goals (as the case requires), without necessarily having to use all of them.

| N | Objective | Description |
|---|--|---|
| 1 | apply knowledge | Allow participants to apply relevant course knowledge in practice through decision-making and debate |
| 2 | instantly usable interfaces for 20-30 | Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use |
| 3 | one teacher | Allow a single teacher to facilitate and debrief the activity, using an ecology of displays |
| 4 | reflect on assumptions via decisions | Encourage participants to reflect on assumptions through decision making and discussion |
| 5 | refresh by moving and talking | Allow participants to refresh from exhausting lectures by moving and talking |
| 6 | team incentives for negotiation | Encourage negotiation by letting participants follow different goals but requiring the consent of teammates |
| 7 | social mechanisms and play | Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play |
| 8 | ambient visualisations for awareness | Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data |
| 9 | inclusion and equitable accessibility | Allow all participants to contribute equitably to the activity and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics |

Table 8.2.: The 9 Gameplay Objectives

8.2.1.5. Three Bubbles

The Three Bubbles model (Figure 8.6) applies Yuill & Rogers' (2012) *mechanisms for collaboration* to three hierarchical layers of group size: class, team and region. By suggesting how these mechanisms can be meaningfully constrained in a whole-class context, the model aims to support design, orchestration and evaluation. A detailed rationale is provided in Chapter 7, Section 7.16.

The model's suggestions for design can be summarised as follows: The class should be divided into teams and teams into regions. By sharing a single device, each region controls a local part of the simulation. To encourage collaboration between regions, the activity should involve subtle incentives for regions in the same team to prioritise the wider team's long-term interests over their own, regional short-term interests. For instance, if the activity is a simulation then incentives might include sharing a team-wide resource across multiple rounds of decision making.

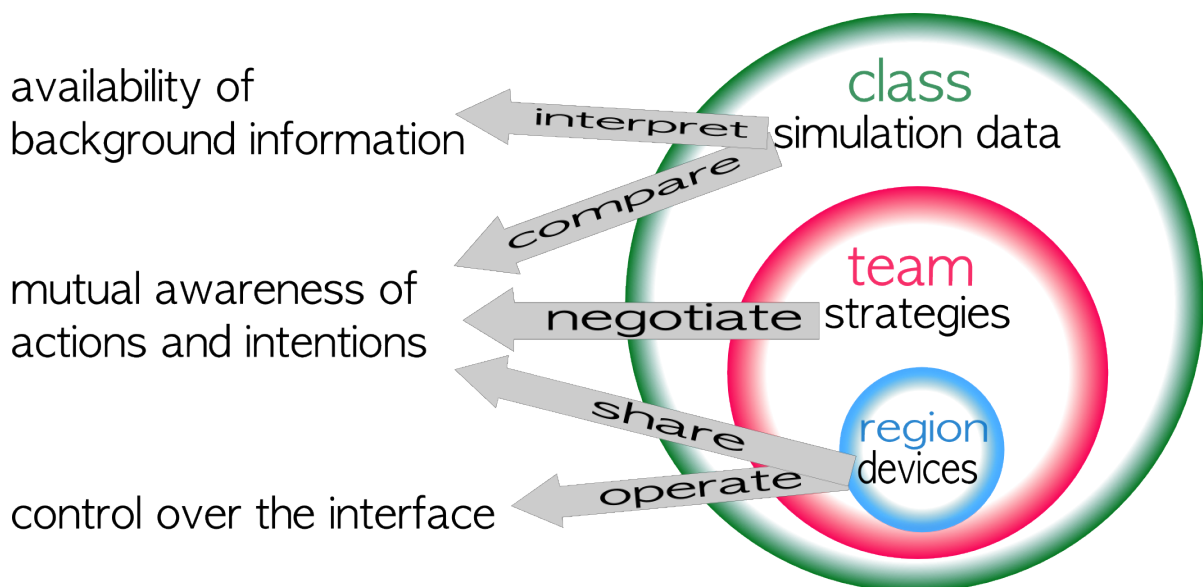


Figure 8.6.: Three Bubbles: mapping mechanisms to different scales of social / physical space

To operationalise the model during evaluation, two methods are proposed to direct the analytical focus to relevant indicators of successful collaboration. First, a simple *rule of thumb* suggests that “regions should share, teams should be aware, and the class should compare”. Secondly, more in-depth evaluation can be guided by the following 5 questions that correspond to the 5 arrows in the figure above:

1. (How) does operating its own device provide *each region* with control over the interface?

2. (How) does sharing the devices support mutual awareness of (a) actions and (b) intentions *within each region-bubble*?
3. (How) does negotiating strategies support mutual awareness of (a) actions and (b) intentions *within each team-bubble*?
4. (How) does comparing simulation data support mutual awareness of (a) actions and (b) intentions *within the whole class-bubble*?
5. (How) does interpreting simulation data make background information available *for the whole class*?

Based on the assumption that all three mechanisms (left in the figure) are essential to successful collaboration, it is worth highlighting in the figure that the first and third mechanism have only one *arrow* pointing to them, whereas three arrows address “mutual awareness”. The latter fact emphasises that mutual awareness may be managed in various (alternative or additional) ways. It even allows removing the *team* bubble entirely, while still providing all the essential mechanisms. Despite a certain flexibility regarding how to manage mutual awareness, the model nevertheless insists on two core principles:

1. private control for the regions;
2. public information for the class.

Finally, whether the 5 *arrows* should occur at the same time or at different times is not prescribed by the model. This is worth pointing out, as it highlights that Three Bubbles is compatible with Multi-Plane Groupwork as well as Single-Plane Groupwork (see below).

8.2.1.6. Multi-Plane Groupwork

Multi-Plane Groupwork (MPG) describes the idea that collaboration may take place simultaneously at different levels (“planes”) of group size. For example, pairs may be sharing a display together whilst simultaneously interacting with other pairs to coordinate their actions towards a larger team effort. In this example collaboration takes place at two planes: *pair* and *team*. The emphasis is here on *simultaneously*. This is in contrast to common approaches - which I call Single-Plane Groupwork (SPG) - that conceptualise pair work, groupwork, whole-class sharing as temporally separate phases of an activity (e.g. the jigsaw model).

Figure 8.7 shows an example of an MPG task within a larger activity. The planes in this example are arbitrary. In principle, MPG could involve any number of simultaneous planes, including pair, small group, team, classroom, school, community, world, etc., in accordance with Dillenbourg & Jermann's (2010) notion of plane.

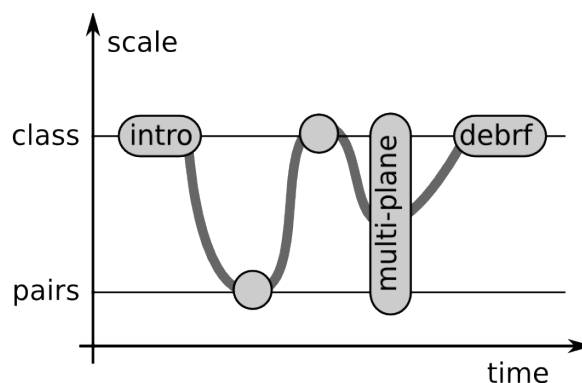


Figure 8.7.: Example CSCL script: Joint introduction, switching between whole-class and pair tasks, followed by a Multi-Plane Groupwork task and joint debriefing.

When deciding between MPG or SPG, designers should be aware of a trade-off of benefits. Whereas MPG may be beneficial in some cases for supporting peer social motivation and control (see Chapter 7, Section 7.18.0.1), SPG may be preferable when the group task requires intense focus.

8.2.1.7. Spatial Formations

Certain spatial configurations between (adult) peers and technology were found more conducive to inclusion and equitable participation than others (see Chapter 7, Section 7.12). The following two tables provide an overview of effects that were commonly found in regions (Table 8.3) and teams (Table 8.4).

| Region Formation | Implications for accessibility |
|--------------------|--|
| Single driver | individual control |
| Pair | equitable sharing |
| Three in a row | equitable sharing |
| Four in a row | equitable sharing |
| Symmetrical corner | equitable sharing |
| Corner-L | consistently observed in association with physical exclusion |

Table 8.3.: Region formations and their commonly observed implications

| Team Formation | Access to Control | Access to Information | Awareness |
|---------------------------|--|-------------------------------|---|
| Panopticake | equitable | high for navigators | high |
| Circle seating | equitable | high | high |
| Long table seating | encourages problematic Corner-L formations | high | high within regions, low across table |
| Row of drivers | individual | unequal due to standing crowd | high between neighbours, low across distance |
| Driverless row of tablets | unequal, confusing | unequal due to standing crowd | unequal due to standing crowd (visual and physical obstruction) |

Table 8.4.: Team formations and their commonly observed implications

8.2.1.8. Benefits of Rule-Changing

According to the analysis in Chapter 6, collaborative rule-changing can encourage participants' constructive reasoning about simulation parameters and assumptions. The analysis suggests that there are 7 aspects to how an activity is transformed when participants are allowed to change simulation rules collaboratively:

1. The rules are scrutinised collaboratively, rather than individually.
2. Critique is framed in discussion, rather than complaints.
3. Peers are consulted in addition to the teacher.
4. Reflection focuses on assumptions, rather than punishment and reward.
5. Rules are actively explored, rather than taken for granted.
6. Groups immerse themselves more deeply in the activity.
7. Focus shifts from pure gameplay to modelling and game design.

An alternative representation of this list is shown in Figure 8.8. Here the seven aspects are mapped horizontally to parts of a sentence. This is done twice, with the upper sentence describing a less desirable case, and the lower sentence describing a more desirable case, respectively. The two sentences can be read as *before and after* and should be seen as extreme examples. The purpose of this format is to highlight the polarities in a compact way.

Both representations are potentially useful during evaluation. They provide an easy-to-use checklist for researchers and teachers to assess high-level learning in the context of collaborative activities that involve rule-changing.

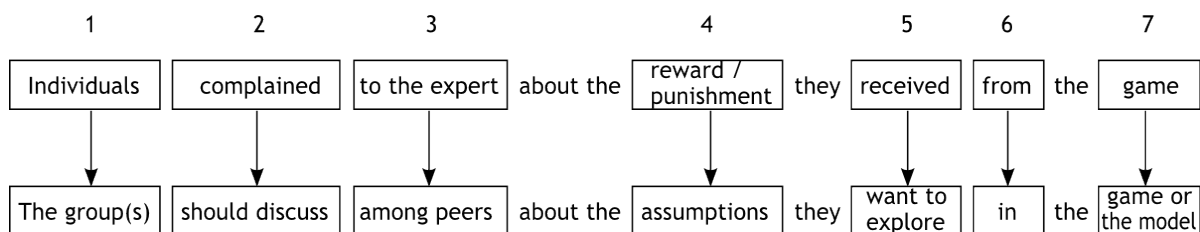


Figure 8.8.: The 7 Rule-Changing Benefits

8.2.2. Recommendations

The following tables list a variety of recommendations. An attempt was made to group the recommendations into categories (or types). This grouping is somewhat arbitrary - other groupings are also possible - and is mainly for the sake of making the whole list easier to read. The categories are titled *Orchestration strategy*, *Game design*, *Rule changing*, *Region interface* and *Ambient display*.

8.2.2.1. Orchestration strategy

The recommendations in Table 8.5 concern the overall design and realtime management of a lesson.

| Recommendation | Description |
|-----------------------------|---|
| Appropriate prior lectures | Prior lectures should establish basic conceptual understanding at an adequate depth for participants to make sense of the simulation. |
| Teacher-controlled stepping | Give the direct control over moving the activity to the next phase |
| Practice match | Allow groups to familiarise themselves with the controls and information without the pressure/complexity of winning conditions |
| Surfacing | Use the activity to help the group raise latent beliefs and make assumptions explicit. |

Table 8.5.: Recommended orchestration strategies

8.2.2.2. Game design

The recommendations in Table 8.6 concern aspects of game design that were found beneficial.

| Recommendation | Description |
|------------------------------|--|
| Introduce conflict | Encourage debate by giving the groups something to disagree on. |
| Teaching without preaching | A simulation should be a morally neutral tool, aiming to make explicit (rather than manipulate) participants' attitudes towards controversial topics and decisions. |
| Collaborative prioritising | Present the groups with different options to use a limited resource, forcing them to debate priorities |
| Decoupled team play | Decisions made in one team should be independent of the other team. |
| Discrete decision phases | Decision-making should take place in clearly delimited time periods during the activity. |
| Discrete decision types | Design the interface to ask for 'crisp' types of input that groups can be easily describe in words. For instance, a series of boolean or integer values - as opposed to curves or sliders. |
| Decision replacement | Present groups with default decisions that they only need to change, rather than making decisions from scratch. (This applies to gameplay and rule-changing decisions) |
| Easy first choice | Make default values easily recognisable as poor decisions to incentivise initial changing |
| Tragedy of the commons | Design the simulation such that regional short-term interests conflict with a team's long-term interests |
| Performance-independent pace | The pace of progression through a series of tasks should be independent of how groups or individuals perform within each task. Aka "gaming on a rail" |

Table 8.6.: Recommended Game designs

8.2.2.3. Rule-changing

The recommendations in Table 8.7 apply in cases where a game or simulation allows participants to change its rules.

| Recommendation | Description |
|-------------------------------|--|
| Collaborative rule-changing | Allow groups to change the rules for the next match |
| Keep rules simple | Avoid overwhelming the group with overly complicated choices during rule-changing |
| Keep rules relevant | Ensure that all available options within rule-changing are relevant to the learning need |
| Keep rules orchestration-safe | Prevent participants from making rules that could break the orchestration, e.g., require too much time |
| Combined winning conditions | Offer combined winning conditions if required, e.g., by a need for more realism or reflection |
| Large-group tablet sharing | For a short period of time it may be feasible to let up to 15 participants share a tablet screen to make decisions |

Table 8.7.: Recommended rule-changing strategies

8.2.2.4. Region interface

Table 8.8 summarises recommendations for designing the tablet interfaces.

| Recommendation | Description |
|--------------------------|---|
| Plus/minus buttons | Allow participants to make decisions by increasing and decreasing a default values. This is considered preferable, e.g., to sliders because the inherent upper and lower limit might bias participants' perceptions of what is <i>relatively</i> high or low. |
| Large fonts | Allow participants to read devices and displays from a distance or angle |
| Small amount of text | Allow participants to read devices and displays from a distance or angle |
| Local/short term control | The control and information provided on a region's device should be limited to that region and round |
| Large buttons | The buttons (or selectable items) should be large enough for users to hit them without taking much time to aim. |
| Simple button shapes | If image buttons have a complex shape, then it might be advisable to surround them with a simple shape (e.g. rectangle or circle) that responds to touch. |
| Immediate touch feedback | When touched, a button should provide immediate visual feedback, noticeable for other participants who may be watching peripherally. |
| Delayed effects | If appropriate, give the user some time to draw their hand off the tablet screen before selections take effect, in order to avoid occlusion. |

Table 8.8.: Recommendations for designing region interfaces

8.2.2.5. Ambient display

Table 8.9 summarises recommendations for designing the ambient displays.

| Recommendation | Description |
|---------------------------------------|---|
| Ambient history | Provide a persistent display of simulation data from past rounds |
| Everything stays | If no information is ever removed from the display during a session, groups can always refer back to it and reflect |
| Growing data | Add more information after every round |
| Data growth follows reading direction | Add new data in the direction of reading - e.g., at the bottom or on the right in the case of English |
| Causality follows reading direction | E.g. with English, show input data on the left/top and output data on the right/bottom |
| Global/long term information | Ambient information should be persistent and global across teams and regions |

Table 8.9.: Recommendations for designing ambient displays

8.3. Conclusion

This chapter presented the CLOVER framework as a summary of previous findings expressed in abstract terms. A distinction was made between *tools* (models for high-level conceptualising) and *recommendations* (low-level themes to support individual decisions). A graphical overview of the tools was developed to allow design teams to see all the figures at a glance. Since the tools were already described at a reasonable degree of abstraction earlier in this thesis, summarising them and adding cross-references was considered an adequate form of presentation. By contrast, the recommendations were named, described and categorised specifically for the purpose of presenting them in this chapter. The next chapter is dedicated to evaluating the usability and usefulness of CLOVER in other design contexts.

Chapter 9.

Validating the framework in the wild

9.1. Overview

This chapter investigates to what extent the CLOVER framework can be applied in other relevant cases. The question is whether the proposed abstractions (i.e., tools and recommendations) are useful for different designers, teachers and researchers when developing and evaluating sustainable whole-class activities. Designing for sustainability in the wild implies (a) handing over control to the teacher and (b) using available resources that the teacher can manage in real time. Two new activities are described that were designed using CLOVER and use tablets and a single projector to encourage discussion in classrooms. The first activity (UniPad) is a whole-class simulation that was closely modelled after 4Decades and aims to help pre-university students discuss personal finance decision in classrooms. The second activity (ComfyBirds) aims to transform an existing desktop application into an orchestrable whole-class activity. Rather than being a simulation, ComfyBirds allows pairs of children to discuss ambiguous words in a series of joking riddles, thus testing CLOVER’s adaptability to a wider range of pedagogical goals. Both applications were found orchestrable by teachers without requiring excessive preparation time, maintenance or technical skills. The analysis shows how the tools and some of the recommendations from the CLOVER suite were put to use in iterative design and evaluation, thus providing support for their broader applicability. In addition, a number of unexpected requirements were revealed in the design process that contribute to a better understanding of what it means to design for sustainability in a whole-classroom context. Figure 9.1 summarises the main research question, data sources and analysis methods used in this chapter.

| | chapter 4 | chapters 5-7 | chapter 8 | chapter 9 |
|-------------------------|--|--|--|--|
| Question | <p>Q1</p> <p>How can the whole group engage with the topic?</p> <p>a) What types of engagement goals are desirable and feasible?</p> <p>b) What mechanisms can effectively support these goals?</p> | <p>Q2</p> <p>How should one orchestrate the activity?</p> <p>a) What constitutes the orchestration load?</p> <p>b) How should the orchestration load be balanced between teacher and technology to achieve feasibility?</p> | <p>Q3</p> <p>What are useful abstractions to aid future design, orchestration and evaluation?</p> <p>a) How can we generically describe the factors that make or break the design?</p> <p>b) What are practical recommendations for teachers and designers?</p> | <p>Q4</p> <p>Are these abstractions also useful when designing for sustainability?</p> <p>a) Is the CLOVER framework applicable in different settings?</p> <p>b) What design factors are critical for achieving sustainability in the wild?</p> |
| Study | 4Decades (lab) | 4Decades in the wild | | UniPad Comfy Birds |
| Data | lab video, field notes system logs teacher interviews student questionnaires debriefing transcripts | in-situ video, field notes system logs, screenshots teacher interviews student questionnaires debriefing transcripts | findings from 4Decades and literature | notes from design meetings with teachers teacher interviews field notes system logs |
| Analysis methods | lab-based prototyping qualitative analysis | detailed video analysis qualitative analysis of system screenshots | reflection on design development of a conceptual framework | lab based prototyping preliminary feasibility evaluations in the wild qualitative analysis |

Figure 9.1.: This chapter in the context of other chapters.

9.2. Questions and aims

The study was guided by the following main question.

■ **Q4.** Can CLOVER support sustainable classroom activities?

This question implies two sub-questions.

- **Q4-A.** Can designers, teachers and researchers use the CLOVER abstractions in concrete situations of design, orchestration and evaluation?
- **Q4-B.** What are salient design factors for sustainability?

The following aims were addressed.

1. to design and evaluate two whole-classroom activities (for verbal engagement and reflection) that are feasible and sustainable under real conditions (design studies 1&2);
2. to see if teachers can use the CLOVER abstractions during collaborative design (design study 1);
3. to investigate whether CLOVER can help to replicate relevant findings from 4Decades in a new setting (design study 1);
4. to explore how far the CLOVER abstractions can be extended to support different types of activities (design study 2);

Both studies aimed to use CLOVER to inform novel design decisions and test the designs in real settings. However, the two studies were complementary insofar as they took different approaches to the overall goal of validation. Whereas the first study closely followed the premises of CLOVER aiming to replicate the findings of 4Decades with a different topic and target group, the second study deliberately started from a different set of pedagogical goals, in order to test how far the abstractions in CLOVER can be extended outside of their original context of building simulations. These two complementary approaches are used to illuminate CLOVER from complementary perspectives and to validate complementary aspects.

9.3. Chapter structure

The following sections provide an overview of the methodological approach and outcomes of the two validation studies. The studies are then described in detail (the UniPad study and the ComfyBirds study). The chapter ends with a discussion section and summary of conclusions.

9.4. Method

Two studies are described in this chapter. Each study focuses on the design of a particular whole-class activity and aims to assess the design regarding three aspects: (1) the activity's overall feasibility in real classrooms, (2) the activity's sustainability and (3) the usability and usefulness of CLOVER in the design and evaluation process. Different methods, data and analytical foci were used in addressing each of the aims.

Similar to 4Decades, both studies took an iterative design approach, involving teachers and other researchers as domain experts and members of the design team. These people contributed actively by (a) bringing to the table a holistic understanding of students' learning needs and relevant classroom constraints, (b) facilitating the designed activities and (c) providing expert feedback on successes and failures during the sessions regarding student engagement, orchestrability and sustainability.

9.4.1. Assessing feasibility

A range of feasibility-related factors were investigated, including the application's usability, equitable participation and the teacher's actions. In contrast to the 4Decades study, which had involved months of detailed qualitative video and questionnaire analysis, the focus here was on lean evaluation methods, using mostly field observation and teacher interviews. Relevant CLOVER abstractions were used with the goal to focus and accelerate evaluation of whole-class activities. In addition, one UniPad classroom session was described in detail from the teacher's perspective, using descriptive video analysis.

9.4.2. Assessing sustainability

Sustainability and scalability are closely related, according to Dillenbourg’s (2012) suggestion that “If we design for heroes, we lose scalability (there are few heroes) and sustainability (heroes get tired)”. Following this conjecture, we could arguably assess sustainability and scalability by asking *how much of a hero does it take?*

As a practical rule of thumb that comprises relevant factors, I propose the “Hero Quotient” which can be defined as the amount of required preparation time, maintenance and orchestration load in relation to a typical teacher’s daily capacities of time and energy (Figure 9.2). In principle, the higher the value, the more of a hero is required for conducting a particular activity in a particular setting.

$$\text{hero quotient} = \frac{\text{preparation time} + \text{orchestration load} + \text{maintenance load}}{\text{typical teacher's daily capacities}}$$

Figure 9.2.: Formula for determining the hero quotient

Quantifying each of the variables is not straightforward, since they may vary widely depending on personal and contextual circumstances. These concerns necessarily limit the degree to which the sustainability of a given application can be measured in objective terms. In this sense, the “hero quotient” should not be seen as an actual mathematical equation but rather a concise summary of relevant constraints to keep in mind. Describing these constraints and their implications from the perspective of a given case is arguably the most useful assessment of sustainability we can provide in this context - short of observing actual, sustained use.

9.4.3. Assessing the use of CLOVER

Both studies furthermore aimed to assess whether the CLOVER abstractions are found usable and useful by teachers in concrete design situations. Meeting protocols and interviews with the teachers were the primary source of data in this regard, in addition to the actual sketches and prototypes. Whereas the UniPad study aimed to replicate findings from 4Decades in a new setting, ComfyBirds aimed to test CLOVER’s adaptability to different design premises.

9.5. Overview of validation studies

This section provides an overview of the goals and key outcomes of the two studies (UniPad and ComfyBirds) with regard to validating CLOVER.

1. **UniPad** is a whole-classroom simulation that was designed to help participants make personal spending decisions under a limited budget, encouraging them to prioritise their needs and wants (Kreitmayer et al., 2013). The design was a direct successor to 4Decades and in some ways can be thought of as a miniature version, since it operates on a smaller time scale (15-25 minutes for the whole activity) and uses fewer resources (4 tablets and one projector), whilst addressing the same essential learning need, i.e., for groups to discuss and prioritise real-life future spending decisions. Moreover, UniPad inherits from its predecessor many key interaction principles, including the general approach to simulation-based teaching, the structure of Three aubbles and a similar approach to interface design.
2. **ComfyBirds** is a whole-class activity to increase children's reading comprehension by helping pairs discuss language ambiguity in joking riddles. This learning goal focuses on verbal engagement and reflection. However, the task differs markedly from 4Decades and UniPad, by (1) involving a set of "correct" solutions and (2) relying on the teacher to provide intense scaffolding to help children understand the rationale behind these solutions. Based on an existing desktop application for single pairs (Yuill 2007) that had previously been shown to increase children's reading comprehension (Yuill 2009), CLOVER was used to scale the design up from a single-pair task to a whole-classroom activity. In contrast to UniPad and 4Decades, ComfyBirds is not a simulation (although one can call it a game). This different use of CLOVER was intended for exploring how far the abstractions can be extended outside their original context of building simulations based on domain models.

In short, two complementary approaches to testing were taken in the two studies. With UniPad the goal was to let CLOVER's proposed tools and recommendations guide the design. By contrast, ComfyBirds aimed to test CLOVER's boundaries with an edge case, to see where it breaks and how.

As Figure 9.3 shows, UniPad validates 7 of the 8 CLOVER tools, i.e. all except the Benefits of Rule-Changing. The latter was not needed in this case, since the activity eventually had to fit into short workshop slot, allowing only one single match to be

played. Apart from this minor limitation, UniPad made use of all the CLOVER tools in the design and some were used in the evaluation.

In contrast, ComfyBirds provided validation for the *Triple-i Hierarchy* and some of the key aspects of *Gameplay Objectives*, *Three Bubbles* and *Spatial Formations*. Validation of the latter three was limited by the fact that ComfyBirds did not use teams (at least not in the current early version - teams are considered as a future design option). Three of the 8 tools (namely *Model Sim Game*, *Lecture Sim Discussion* and Benefits of Rule-Changing) were not applicable, since ComfyBirds was not designed as a simulation. *Multi-Plane Groupwork* was deliberately not used, as the nature of the task suggested having dedicated phases for the children to focus in pairs. A seated layout was chosen with pairs facing the teacher. Although pair use was found essential for sharing decisions (as advocated by *Spatial Formations*) initial pilot studies with young children - who liked to run around - raised questions about how to best keep the tablet drivers in their driver's seats.

| # | Tool name | Follow-up study 1 UniPad | Follow-up study 2 Comfy Birds |
|---|-------------------------------|---|---|
| 1 | Model Sim Game | Initially built with a formal winning condition (score). Later decided to removed it. Scenarios help adapt the activity to other groups | (not a simulation) |
| 2 | Lecture Sim Discussion | Simulation was used post lecture Followed by debriefing | (not a simulation) |
| 3 | Triple-i Hierarchy | Compatible with the teachers' focus on pedagogy during design | VALIDATED |
| 4 | Gameplay Objectives | Generally achieved + highlighted new insights about when to use Multi-Plane versus Single-Plane Groupwork | VALIDATED Except #6 and #7 (no teams currently) |
| 5 | Three Bubbles | Guided the design. Rule of Thumb very helpful for Quick assessment of an activity's overall feasibility and inclusivity. | ADAPTED Sharing in pairs and Comparing as a whole class. Future option to use teams. |
| 6 | Multi-Plane Groupwork | Guided the design but not adopted by groups. New insights and questions re needs, benefits, feasibility and potential tradeoffs | Designed as Single-Plane, switching between pairs and class, allowing pairs to focus. Preliminary lab test showed that children might want Multi-Plane. |
| 7 | Spatial Formations | Used in discussion of findings and future design potential | IN PROGRESS Pilot study validated Sharing in pairs |
| 8 | Rule-Changing Facets | FEATURE IN PREPARATION | (not a simulation) |

Figure 9.3.: Overview of the two validation studies

In addition to the *tools*, the following CLOVER *recommendations* were applied and validated in both studies (Table 9.1).

| Name | Description | Type |
|------------------------------|--|------------------------|
| Teacher-controlled stepping | The teacher should have direct control over moving the activity to the next phase. | Orchestration strategy |
| Performance-independent pace | The pace of progression through a series of tasks should be independent of how groups or individuals perform within each task. | Orchestration strategy |
| Decision replacement | Present groups with default decisions that they only need to change, rather than making decisions from scratch. (This applies to gameplay and rule-changing decisions) | Game design |
| Global/long term information | Ambient information should be persistent and global across teams and regions | Ambient display |
| Ambient history | Provide a persistent display of simulation data from past rounds. | Ambient display |
| Local/short term control | A region's device should provide information and control specific to that region and round | Region interface |
| Large fonts | Allow participants to read devices and displays from a distance or angle. | Region interface |
| Small amount of text | Allow participants to read devices and displays from a distance or angle. | Region interface |
| Large buttons | The buttons (or selectable items) should be large enough for users to hit them without taking much time to aim. | Region interface |
| Simple button shapes | If image buttons have a complex shape, then it might be advisable to surround them with a simple shape (e.g. rectangle or circle) that responds to touch. | Region interface |
| Immediate touch feedback | When touched, a button should provide immediate visual feedback, noticeable for other participants who may be watching peripherally. | Region interface |
| Delayed effects | If appropriate, give the user some time to draw their hand off the tablet screen before selections take effect, in order to avoid occlusion. | Region interface |

Table 9.1.: CLOVER recommendations that were applied in both studies

The following CLOVER recommendations were applied only in UniPad (Table 9.2).

| Name | Description | Type |
|---------------------|---|------------------------|
| Surfacing | Use the activity to make latent beliefs and assumptions explicit. | Orchestration strategy |
| Decoupled team play | Decisions made in one team should not influence the other team. | Game design |
| Easy first choice | Make default values easily recognisable as poor decisions to incentivise initial changing | Game design |

Table 9.2.: CLOVER recommendations that were applied in UniPad (but not in ComfyBirds)

In the following I will report on both two studies in detail. Each study is described with regard to aims, method, setting, approach to design and evaluation, and a detailed description of the application. Findings are presented and discussed in regard to feasibility, sustainability and use of the CLOVER framework.

9.6. The UniPad study

Similar to 4Decades, the UniPad simulation was designed in a training course context, aiming to facilitate large groups' discussions of real-life challenges in ways that are socially engaging, inclusive and orchestrable for a teacher. By using the CLOVER framework to guide a new design, the study aimed to replicate findings from 4Decades, thus empirically validating CLOVER's ability to support sustainable designs for different user groups and subject domains. The topic of student finance was chosen and iterative, design was conducted in collaboration with London-based charity MyBnk (www.mybnk.org) who provide finance training workshops in school classrooms. Some of the workshop organisers had seen a presentation of 4Decades and suggested that a similar activity might be designed for pre-university students *"to make budgeting more interactive"*. Below I describe the context of the study, lab-based design iterations and findings from two classroom sessions where different teachers conducted the UniPad simulation in a workshop context.

9.6.1. Background

Student debt is a major problem in the UK and other countries, partly due to rising tuition fees and a lack of finance education in schools. Teacher-led group activities have mostly relied on paper-based budgeting tasks, sometimes involving pocket calculators, card games or board games (O'Neill 2008). Although recent studies (besides commercial offers such as www.bamzonia.com) have begun to explore the use of digital games (Jones and Chang 2012) and virtual worlds (Johnson 2012) to support finance literacy, these applications were designed for single users with mouse and keyboard, rather than taking the face-to-face benefits and limitations of classrooms into account.

9.6.2. Aim of this study

In addition to the overall aims of this chapter (see Section 9.2), the UniPad study specifically aimed to validate CLOVER by applying the simulation-based approach in a new setting, addressing a different target group and collaborating with different teachers.

9.6.3. Setting

The UniPad study was conducted in the context of a series of finance training workshops aimed at students. The explicit overall goal of the training workshops was to increase students' knowledge, skills and confidence regarding sensible spending decisions. To address this goal, lectures and various exercises encouraged participants to reflect on their needs and wants, what their priorities are, and how those could be reconciled with a limited budget. In principle, this learning goal was similar to the goal of 4Decades, which had also been about prioritising spending.

9.6.4. Approach to design and evaluation

Iterative design was conducted in collaboration with MyBnk staff, notably the head of education, head of innovation and strategy and several trainers (from now on referred to as teachers, for consistency). Due to their domain expertise, extensive workshop experience and understanding of the target group's learning needs, the teachers were competent advisors regarding pedagogical questions. Pedagogical aspects were considered paramount to all other design concerns, in accordance with the Triple-i hierarchy.

Teachers were involved in all design iterations. Particularly, this included discussing sketches and proposed features together as well as planning, conducting and reflecting on user tests. Moreover, MyBnk's responsibility was to provide the domain-specific material for the simulation (in analogy to Stephen Peake providing the climate model to 4Decades). This material included text, numbers and graphics which I integrated into the software, being the only programmer on the project.

Some of the early lab tests were directed by myself. Those were essentially technical dry runs with groups of teachers pretending to be students. The aims of these early tests were to (a) test the stability of the application, (b) familiarise teachers with the setup procedures, (c) collect feedback and ideas from teachers on the instructional design.

Later tests were conducted by MyBnk on their own, using stable prototypes. Being able to set up and use the simulation independently allowed MyBnk to conduct their usual focus group tests. One step before classroom deployment, MyBnk tested the activity once in a community centre, in order to practice the setup and facilitation procedures under more "wild" conditions (i.e., similar to classrooms but without the time constraints).

Finally, two sessions were conducted in real classrooms with the intended target audience, i.e., pre-university teenagers (15-17 years old). Different methods were used in the two sessions. While the first session's evaluation relied on field notes and a post interview with the teacher, the second session used video analysis for a detailed description of the teacher's actions in during the simulation. The rationale for using diverse methods was to gain experience with different approaches to applying CLOVER tools in evaluation. Furthermore, both sessions investigated the application's feasibility and sustainability.

9.6.5. The application

The UniPad simulation represents 12 months in the lives of four university students ("characters" from now on) whose only source of income is a student loan. Each month they face a variety of spending decisions, including basic living expenses (rent, bills and food), tools, gadgets, clothes, fun activities, seasonal events and emergencies. Living in shared apartments together, they can save some money by buying only *one* television (or toaster, blender, vacuum cleaner, etc respectively) per apartment, rather than one per student. This requires the characters to negotiate who buys what. Each of the four characters is controlled by a group of participants (pre-university students) that share a simple tablet interface. Their virtual bank balances, purchased items and special events

are visualised on an ambient display. The latter is also used by the teacher who controls the pace of the activity using a handheld interface.

9.6.5.1. Procedures

Figure 9.4 details the overall flow of the activity, broken down into essential procedures. First the teacher sets up the tablets and projector by directing the web browser to UniPad's URL (typically by typing the URL into the classroom's existing PC that drives the projector, while the tablets offer prepared bookmarks). The activity starts with the teacher handing out the tablets and allowing each region to enter a name for their character. The teacher then explains the contents and purpose of the interface and starts the first round. The teacher then explains the contents and purpose of the interface and starts the first round.

During the round, ambient information updates instantaneously as the regions change their decisions on the tablets. The teacher closes the round after giving the regions brief notice to finish their decisions. Once the round is closed, the tablets become non-interactive, allowing the teacher to talk to the class and helping the regions compare simulation data with other regions, via the ambient display. After 12 rounds the teacher leads a debriefing discussion, Q&A or lecture as appropriate.

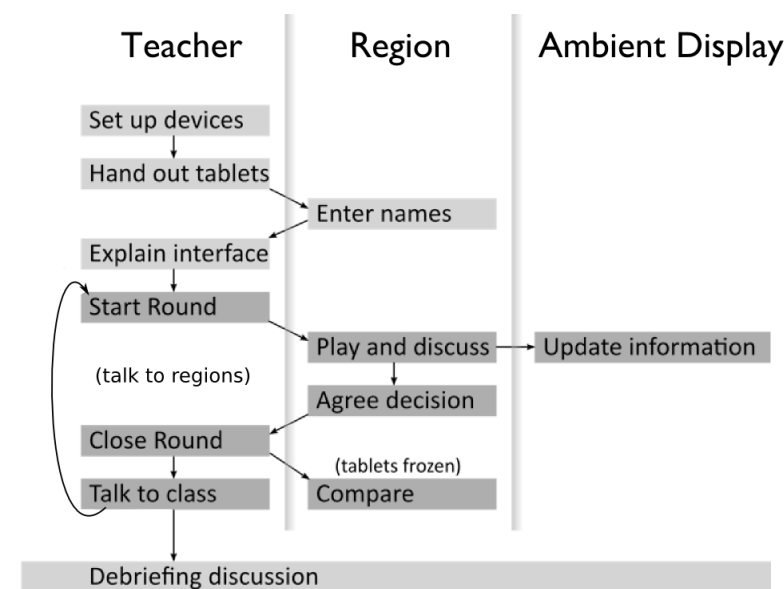


Figure 9.4.: Overview of the procedures

9.6.5.2. Teacher interface

A simple handheld interface allows the teacher to step through the application using a single “PROCEED” button. Figure 9.5 shows screenshots of the interface during a round (left) and after a round (centre). An older version is shown in contrast (right) that had two buttons, one for starting and one for closing rounds. An additional “MENU” button provides extra functionality, such as starting an entirely new match. The interface provides the teacher with essential information:

- Stage of the activity, e.g. “In round”, “End of round”, etc;
- Round number, e.g. 8 out of 12, and name of the month, e.g. April;
- A reminder of what to do at the current stage of the activity, e.g. “Let students make decisions”.

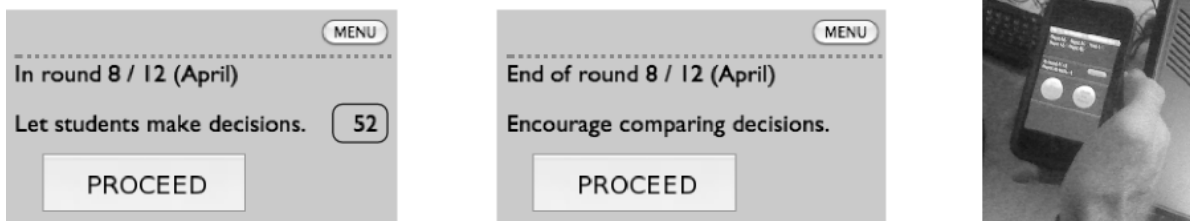


Figure 9.5.: Teacher interface to step through the application

9.6.5.3. Tablet interface

Figure 9.6 shows the tablet interface during the second round (October, see top right corner). The top row shows the name of the character (blacked out) and the flat (A or B). The interface comprises three main columns. The left column shows current possessions, bank balance, income in the current month and an overview of current outgoings. The outgoings comprise monthly payments (such as broadband, phone contracts, etc) and the cost of the selected purchase items.

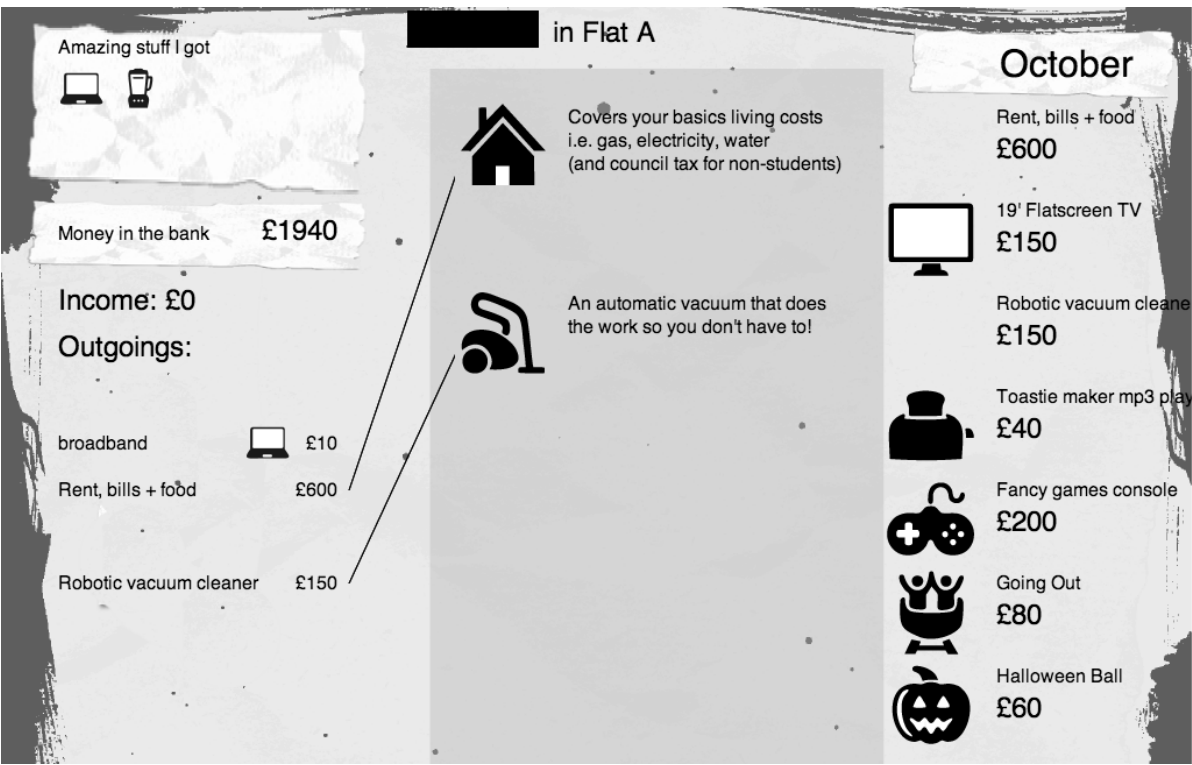


Figure 9.6.: UniPad Tablet interface

The central column shows items that are currently selected for purchase. For each selected item a short text description is shown, providing relevant information, including follow-up costs (e.g. TV subscription). Moreover, the connection between selected items and outgoings is highlighted, making it easier to see all the costs at a glance.

9.6.5.4. Ambient display

Figure 9.7 shows a screenshot of the ambient display, taken from one of the classroom evaluations. The top row shows the four characters' names and bank balances (names blacked out). Underneath the apartments are shown graphically as an interior with icons representing purchased items (e.g. television, toaster, headphones) and activities (photos on the wall). The ambient display was programmed to update the icons instantaneously, as items were selected and deselected for purchase.

A relevant subset of those items also appears in the vertical timeline (bottom half). This month-by-month history shows items in black if they have been purchased, or grey otherwise. The columns represent, from left to right, whether each character has paid their rent (house icons), emergency payments (warning triangle), items that are shared among both characters (TV, vacuum cleaner, toaster, blender, games console, cleaning materials) and finally the pound sign indicating positive (black) or negative (grey) bank balance for each character.

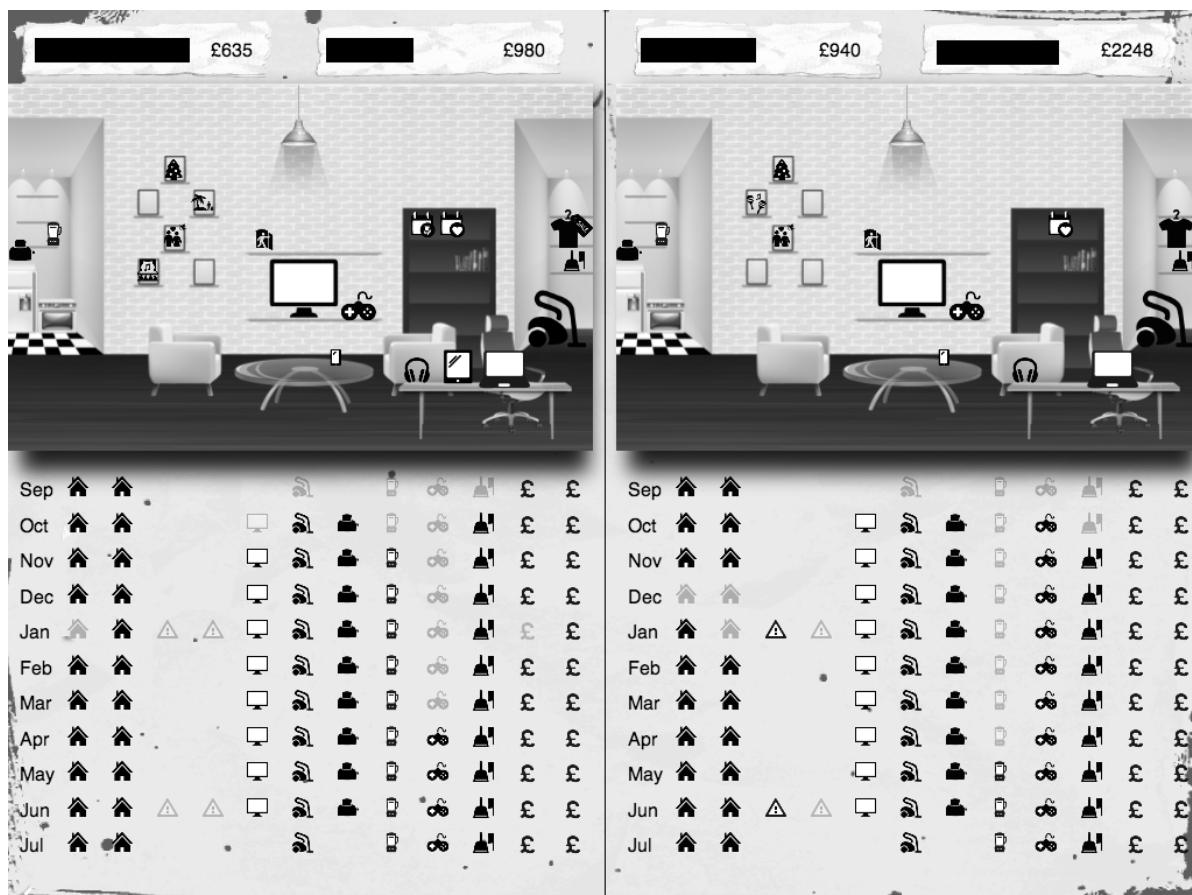


Figure 9.7.: Ambient display, showing both teams' decisions in the next-to-last round (July)

9.6.6. CLOVER-based design process

The collaborative design of UniPad with finance teachers was markedly informed by the CLOVER tools and recommendations. In the following I explain how some of these abstractions helped to (a) conceptualise the educational goals and priorities of the design, (b) support effective forms of communication within the design team that were focused on addressing these goals and (c) provide guidance in concrete design decisions.

9.6.6.1. Educational goals and design priorities

The design goals were focused on a real educational need, i.e., to support personal budgeting decisions classroom settings. Developing effective ways to address this need within their training workshops was MyBnk's primary motivation for taking part in this design. Their workshops were designed and facilitated with a strong emphasis on peer social interaction, rather than pure lectures. One teacher phrased this in an interview as *"getting them to talk is why we're there"*.

It was in this regard that MyBnk's teachers saw potential in leveraging technology. They did not see tablets as a gimmick to enhance individual tasks with fancy graphics and multitouch. Moreover, the tablet's novelty as a medium was not perceived as a motivating factor. Based on their daily experiences with urban teenagers, several teachers suggested that this population had passed the peak of initial excitement about tablets as a new medium. *"We [adults] still think it's new but for them it's already normal, like headphones or sneakers. They love it, but it's one of many things. So two years ago they'd see an iPad and go wow - now we're kind of past that."*

Designing technology as a means to address a well-defined pedagogical need is a core principle behind the Triple-i hierarchy. MyBnk's existing focus on specific instructional goals (i1) and existing priority of face-to-face interaction (i2) over interfaces (i3) provided a beneficial environment for using the Triple-i approach.

9.6.6.2. Designing a model together

In contrast to Triple-i, the teachers found Model Sim Game less intuitively accessible. For instance, I remember suggesting in one of the first meetings that *"first we need a suitable domain model that we can then turn into a simulation and make interactive."* The reactions were somewhat baffled, as it turned out that "domain model" was not

a term that the teachers were used to operating with on a daily basis. Although the misunderstanding was easily clarified, it was a revealing lesson early on in the design process, making me aware that some of the concepts and assumptions from 4Decades may transfer more directly and easily to another context than others.

In this particular case the solution was to focus on specifying a concrete *scenario* from which I then abstracted a computational *model*. The scenario comprised an overall storyline including “characters” (students living in shared apartments), a specific source of income (a student loan), a specific set out purchase options (rent, food, gadgets, activities, etc), and a specific time scale (12 months). This process was revealing insofar that it illustrated the difficult challenge of representing things in simple structures without sacrificing realism. Unlike in 4Decades, where this challenge had been tackled by a single domain expert (Stephen Peake), here it was a collaborative challenge. Over the course of several iterations, the simulation content was negotiated between the teachers’ ideas regarding the need for realism and my ideas regarding the need for simplicity.

A significant milestone in this negotiation process was the formulation of a spreadsheet, defining the exact wordings and numbers of the simulation content, particularly the purchase decisions. Like in 4Decades, this spreadsheet was designed by a senior domain expert (MyBnk’s head of education) but here the focus was on raw data, rather than mathematical equations. An excerpt of this spreadsheet is shown in Figure 9.8, showing some of the 21 purchase options, including text descriptions, prices, and payment options.

The list of purchase options continues as follows: Swanky headphones, Fancy games console, Going Out, Going on a date, Halloween Ball, Christmas presents, January sales shopping, Valentines Day, Taking Mum out for Mother’s Day, Booking a holiday, Festivals / concerts, Getting a cleaner in. Moreover, the spreadsheet was annotated with suggestions for different payment schemes and how the options might be distributed across the months.

- “Rows [2-4]: These all require a one off payment (for equipment) and ongoing contract subscription - is it possible for this to be reflected post purchase? I.e. If the purchase is made, in subsequent months that item is greyed out + an ongoing payment gets taken out of their budget?”
- “Row 6: These are optional electrical items - we could mix up the order if we want (i.e. Rather than having the same item two months in a row, spread them accross the 1st + 2nd 6 month periods)”

| Item | Description | Value | Icon | Shared | Payment Options - NB I think |
|--|---|--|---|--------|---|
| Rent, bills + food | Covers your basics living costs i.e. gas, electricity, water (and council tax for non-students) | £600 | House | n | Monthly amount |
| Smartest phone ever | 3G wifi enabled, mp3 playing smartphone + monthly contract | £100 initially + £30 a month in subsequent months | Smartphone | n | The first month costs £100 + £30 pyt, followed by £30/month in subsequent months |
| 19' Flatscreen TV with 500 channels | 19 inch flatscreen TV + monthly cable/sky subscription | £150 initially + £20 a month in subsequent months | Flatscreen TV | y | The first month costs £150 + £20 pyt, followed by £20/month in subsequent months |
| Laptop packed with features + broadband connection | Laptop with superfast processor, 500GB memory + monthly broadband | £300 initially + £30 set up fee (small print) + £10 a month in subsequent months | Slimline laptop | y | The first month costs £300 + £30 set up fee, followed by £10/month in subsequent months |
| Clothes + shoes | Includes summer / winter clothes, trainers, shoes, accesories | £60 | Clothes + trainers / heels | n | Monthly option available (i.e. doesn't get greyed out) |
| Robotic vacuum cleaner | An automatic vacuum that does the work so you don't have to! | £150 | Irobot round vacuum cleaner | y | One off option |
| Mega deal on a new Ipad! | Super discounted latest 16GB Ipad offer | £333 | Ipad | n | One off option |
| Toastie maker that's also a mp3 player speaker | Blast your tunes while you toast your dinner | £40 | Toastie maker with 'sound' coming out of it | y | One off option |
| Smoothie / cocktail blender | Be super healthy and / or the best party host! | £30 | Blender | y | One off option |

Figure 9.8.: Some of the purchase options that constitute the content of the model

9.6.6.3. A game or a simulation?

Occasional discussions came up as to whether the activity should be generally referred to as a “game” or a “simulation”. Both terms would be adequate, according to CLOVER’s *Model Sim Game* taxonomy. During internal design meetings the teachers mostly said “game” and I found myself adopting this habit, mostly because the word is short, which makes it convenient given how frequently it was needed. However, in presentation context the teachers were conscious about choosing the right word, as they explained, because they were thought to raise different expectations. Whereas “game” was associated with fun, “simulation” was considered to sound more pedagogically mature. These different connotations were named as the deciding criteria when introducing the application to test participants, head teachers, potential funding bodies or students in classrooms.

9.6.6.4. Three Bubbles

To replicate the effectiveness of 4Decades, copying its structure was the first educated guess. Consequently, the class was divided into two teams and the teams into regions. As suggested in Three Bubbles, incentives were introduced to encourage team-wide collaboration. Particularly, some purchase options were defined as shared between regions of a team (fifth column in Figure 9.8). For instance, if a region bought a vacuum cleaner or a television then the other regions would not need to buy another one.

Although the abstract concepts of “team” and “region” were found useful in a technical context (i.e., my source code), the concrete notions of “characters” living in “flats” were preferred by teachers in design conversations and during orchestration. Moreover, teachers were occasionally found to use the word “team” ambiguously, referring to either a flat or a character.

9.6.6.5. Gameplay Objectives and recommendations

The rationales for using tablets and ambient displays were notably informed by CLOVER’s Gameplay Objectives. In addition, various recommendations were used, as summarised above in Section 9.5 (Tables 9.1 and 9.2). Since CLOVER makes no explicit suggestions regarding the actual number of displays and devices, the design deliberately started with minimum hardware requirements, with the option to increase these later if found necessary during evaluation. This implied 2 regions per team (i.e., 4 tablets in total) and 1 ambient display.

9.6.7. In-situ feasibility

This section presents findings from two sessions in real classrooms where MyBnk teachers used UniPad with groups of 15-17 year old pre-university students. The evaluation focuses on how the activity was appropriated by two different teachers in two different room settings.

9.6.7.1. Observed spatial conditions

A comparison of the room layouts (Figure 9.9) shows how the activity was adapted to the given conditions without moving the tables. In both cases the group formations were essentially static, as everyone but the teacher was sitting on chairs. Whereas the teacher in session 1 (left in the figure) managed the class from a fixed position in the front, in session 2 there was space for the teacher to walk between regions and sporadically interact with each individual region.

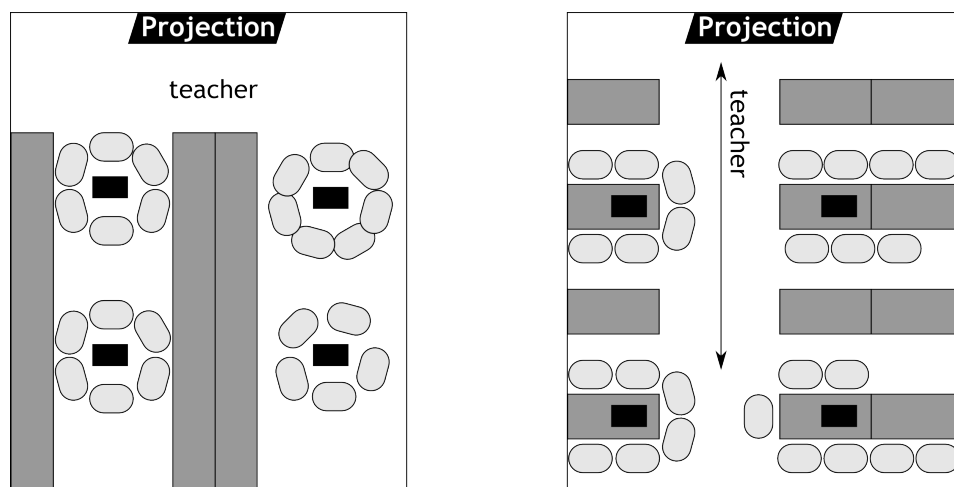


Figure 9.9.: Two sessions showing different types of spacial formations. Left: circular seating without tables. Right: Semi-circles around tables

In both sessions, the regions' particular spatial formations implied that talking between regions was difficult. From their seated positions, none of the participants could easily see another region's tablet, since the tablets were far apart and the view was usually blocked by other participants. No participant stood up during the sessions to walk up to another region. Despite these obstacles, a certain degree of mutual awareness between regions was given by the ambient display. For instance, one student exclaimed, "*Honestly, you're going to uni to play video games? I wanna join you lot, man!*" after realising that another region had just purchased a video gaming console.

9.6.7.2. Classroom session 1

Session 1 (left in the figure) took place in a computer lab with long rows of tables. Two dozen desktop workstations on these tables made it impossible to use any of the spatial layouts suggested by CLOVER. Instead, regions ended up sitting in circles whilst the tablets were held in the middle. This created an obvious conflict of orientation which the groups generally resolved by taking turns holding the tablets. The four regions were 5, 5, 6 and 7 students in size (23 students in total). Despite this large group size, all students were constantly engaged in the activity. The teacher described this in the post interview as remarkable and surprising, contrary to previous experience with group activities:

“I was really surprised. [...] Especially with the - was it 7 or 8? - young boys that were in [anonymised]’s team. The all-boys team. That’s the demographic that I would have to - in another situation - all day I’ve had to marshal them, I’ve had to make sure the boys would stay engaged. Whereas in this, it was more keeping the girls being fair with each other. Because the girls were getting - no we’re not doing that - so they were getting quite catty over the decisions, whereas the boys were - we’re going out, we’re going out. So to have all 7 or 8 of them engaged - for me, certainly, you know, on a day to day basis - when we set them into usually - big groups like that - we play like a shopping challenge. So they go around and they try to compete to that. Usually you got one or two people doing it and the rest just kind of follow them around, pushing each other around and stuff. None of them were doing that. And they were certainly the type to be doing that. So they were very, very engaged.”

Characteristic for this session was that after each round the teacher dedicated some time (typically about a minute) to encourage reflection and discussion among the whole class. For this purpose the teacher made extensive use of the ambient display, each time pointing out the team’s different decisions, encouraging them to compare. Sometimes the teacher asked the class to raise hands. Some examples were, “Who bought their family Christmas presents? Nobody did?” - “Who went on a holiday?” - “Who of you went out?” To the latter question one participant replied “We don’t need to go out, we’ve got smoothies” (after having bought the blender).

Although the tablet interfaces were in ‘locked’ mode between the rounds, the regions were highly engrossed in their regional decisions, so that the teacher occasionally had to calm them down and direct their awareness to the whole class. The teacher addressed this by frequently encouraging social comparison between teams. The final whole-class debriefing was introduced by asking the class which group they thought should be declared as the

winner. Despite the teacher's explicit effort to raise awareness, very little talk happened between the regions. The teacher remarked this in the post interview.

"If I were to play again I would want to build more of a relationship between the people who live together. I didn't get them talking between [regions]. One bought the telly, one bought the games console. Could they have negotiated that? They didn't debate that all. Because, I don't know if they bought it twice. Two games consoles? Presumably, they could have a Playstation AND and Xbox."

As a potential way of encouraging more negotiation between housemates, the teacher suggested potential for a new feature in the application. Specifically, an option was suggested for housemates to negotiate splitting the cost and ownership of shared items.

"Can they have an option where they both split the cost of the TV? And then [debate] the end of the year who gets the TV? I don't know how complicated that would make it. Erm, because we, certainly, as a flat at uni bought a lot of things and then afterwards we were like - well you need to give me the money. And then it was like, well it is used now so it's not worth the same amount."

The only occasion of verbal contact between regions in this session was when the teacher pointed out that one region had failed to pay their rent. This prompted some of their teammates for the other region to turn around, shouting boo and various other expressions of disapproval. This showed that, to some extent, there was a sense of shared identification as a larger team.

9.6.7.3. Classroom session 2

As illustrated in the right half of Figure 9.9 (above), the tables in session 2 were four front-facing rows with a central corridor. The corridor allowed the teacher to walk up to individual regions and exchange brief dialogue. Rather than moving the tables, about half of the 26 students simply turned their chairs around, resulting in groups of 6 and 7 students around each tablet.

Like in the previous session all groups were highly engaged and made efforts to share the tablets equitably. Different strategies were tried, including placing the tablet in the centre or to one side of the table and holding the tablet up vertically towards the peer group. These strategies were successful in allowing everyone to read the screen with reasonable ease most of the time.

The only problematic spatial formation was observed at the bottom right corner of Figure 9.9, where the region was shaped like a capital letter “J”. As the tablet tended to be shared near *heavier* end of this “J-Formation”, one student at the *lighter* end (bottom right corner) was sometimes too far away from the tablet to see conveniently. Nevertheless, the student took part in the discussions and showed no sign of disengagement or frustration that might have prompted the teacher to address the issue (e.g., by asking the student to move over to the other side of the table).

The entire simulation lasted 15 minutes. The teacher’s announcements and actions during the simulation are visualised in the following overview (Figure 9.10). The red and green quotes represent a complete account of the teacher’s announcements to the whole class, with (green) or without (red) simultaneous use of the ambient display. These announcements are represented in full as they were generally uttered in a loud voice and thus clearly intelligible in the audio recording. Yellow quotes, by contrast, represent bits of sporadic dialogue with individual regions. Those were visualised only sporadically, subject to their salience and intelligibility.

The quotes are related by colour and lines to the timeline visualisation. As in Chapter 5, the squares represent relevant behaviours observed in a time grid of 10 seconds, visually distinguished using an arbitrary colour scheme. The bottom rows show the time in minutes and the rounds, i.e., months from September to August. The breaks between the rounds - displayed in white - were generally short, usually shorter than 10 seconds (NB the visualisation rounds up to a 10-second grid). Only the first two breaks were slightly longer, as the teacher explained the content of the ambient display and encouraged the class to reflect and compare between the teams.

The squares in the top row (red) visualise the occurrences of whole-class announcements (“TALK TO CLASS”), including notifying the class of starting and ending round. The second row from the top (“TALK TO TEAM”) is empty, since the teacher consistently focused on addressing either individual regions or the class as a whole. The third row (“TALK TO REGION”, yellow) shows instances when the teacher faced a particular region. These instances happened mostly during the rounds and occasionally involved reading from a tablet (“READ TABLET”, magenta). By comparing red and yellow one can see that the transitions from one round to the next were associated with whole-class focus, whereas engaging with regions happened more during the rounds.

The green squares indicate instances when the teacher used the ambient display (light green: reading, dark green: pointing). While pointing as observed only near the round

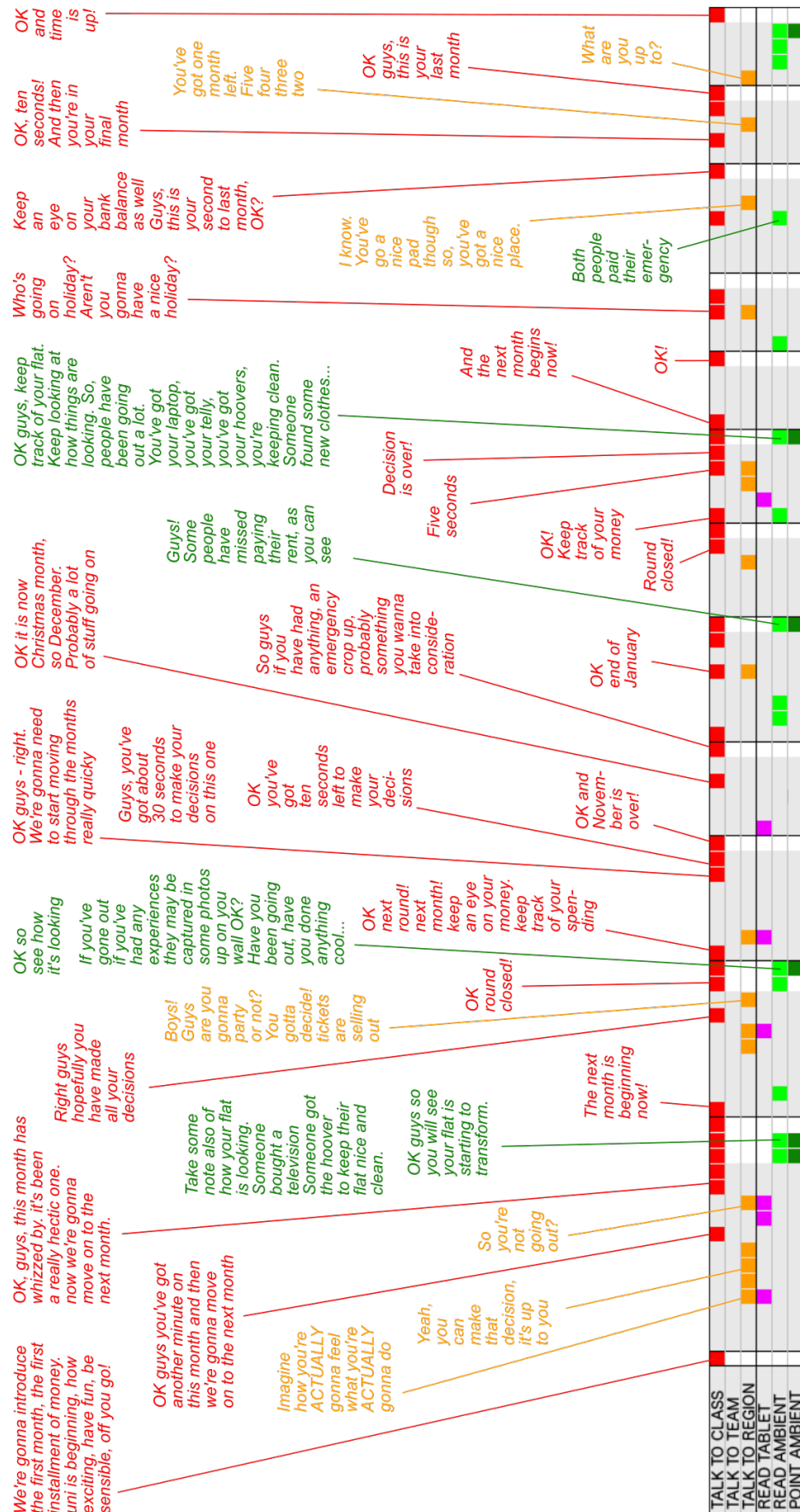


Figure 9.10.: Session 2: Teacher talking to the class (red and green) and regions (yellow)

transitions (5 out of 12 rounds) and always in connection with talking to the class (green quotes), reading also occurred sporadically during the session. At the end of the simulation the teacher called the class' attention and briefly walked them through some relevant observations on the ambient display.

“OK, guys, everyone face the front! Have a look. [...] Most people payed their rent in time [...] At the top you can see your amounts of money that you've got left, so - [name of region], did you guys have any fun? [many excited exclamations and nodding from the region] Oh, you did? But you were very sensible”

The debriefing concluded as the teacher related the simulation to real life situations. This had to be kept brief, due to an earlier delay in the session. Paper-based information material and questionnaires were handed out afterwards.

“Obviously, doing this, it does highlight that the month did go very quickly. You may not have to make that many decisions that quick. But it does highlight some of the decisions that you would have to make. So, sometimes it does boil down to, am I going to pay my rent this month or am I going to go out. You need to, obviously, make sensible decisions around that. Going into minus as well, I did really want to talk to you about banks. There's a lot of information I wanted to get through to you today, but unfortunately time has just crept up. So I am gonna give you a sheet that's got lots of useful website information on [paper is handed out].”

Feasibility in both sessions

Both sessions were completed successfully, with all students taking part from beginning to end. No signs of disengagement or frustration were observed. Rather, both groups were highly energised and focused and participated equitably in the discussion. The observed level of equitable engagement was considered unusual for the age and large number of students per group. The single ambient display was found usable for the teacher and students. Four tablets were found enough to keep a class of 26 engaged. Using the handheld interface, the teachers were able to manage the overall pace and engage with the students in their preferred style. While the teacher in session 1 emphasised social comparison between regions, session 2 was facilitated using a more narrative style. In summary, the Gameplay Objectives were found to be achieved as expected - except for number 6 and 8 - as summarised in Table 9.3 and detailed in the discussion further below.

| N | Gameplay Objective | Comments |
|---|---|--|
| 1 | Allow participants to apply relevant course knowledge in practice through decision-making and debate | achieved |
| 2 | Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use | achieved |
| 3 | Allow a single teacher to facilitate and debrief the activity, using an ecology of displays | achieved |
| 4 | Encourage participants to reflect on assumptions through decision making and discussion | achieved |
| 5 | Allow participants to refresh from exhausting lectures by moving and talking | achieved |
| 6 | Encourage negotiation by letting participants follow different goals but requiring the consent of teammates | negotiation only within regions |
| 7 | Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play | achieved |
| 8 | Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data | awareness of actions but not of intentions |
| 9 | Allow all participants to contribute equitably to the activity and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics | achieved |

Table 9.3.: UniPad satisfied 7 out of 9 Gameplay Objectives.

9.6.7.4. Sustainability

MyBnk teachers were able to set up the activity on their own - including at their own office and in the wild. Since the application was hosted on the web, rather than installed on the portable devices, WIFI access in situ was a requirement. In one location where no WIFI was available, setting up a temporary 3G hotspot using a smartphone was found to be a viable solution.

The minimal hardware set-up (i.e. 4 tablets) was small and light enough to be carried by one person in a bag. In the case of MyBnk this was a real benefit, since many of the teachers relied on public transport to get from one school to another. The additional practicalities were considered by teachers when thinking about possibilities of extending the design. The following excerpt was taken from an interview after the teacher had facilitated a session with 4 iPads and 23 students.

- Researcher: “Would it be better to use more iPads?”
- **Teacher:** *“Apart from the practicalities, I would say more iPads are better.”*
- Researcher: “How many would you suggest, ideally?”
- **Teacher:** *“I think a maximum of 6. Because a class of 30 would be our biggest...so we could have them in groups of 4s or 5s. You don’t need to go smaller than 4s or 5s because they were all integrated. So you won’t need more than 6 iPads. I think you should still keep it 2 houses though. I’d prefer 3 flatmates in 2 houses, rather than 2 flatmates in 3 houses.”*

9.6.8. Discussion

The findings regarding feasibility, sustainability and use of CLOVER are discussed below.

9.6.8.1. Feasibility

As a result of two classroom evaluations, the UniPad simulation was considered feasible in real settings. The design used all of CLOVER’s Gameplay Objectives. The empirical classroom evaluations showed that the design was appropriated as intended, except for objectives 6 and 8, as discussed below.

Objective 1: Allow participants to apply relevant course knowledge in practice through decision-making and debate

This was supported by the observation that all participants were constantly focused on the activity and engaged in the decision making. As promoted by the course lecture, the groups engaged with the notion of needs and wants. Generally the groups aimed to pay for their needs and prioritise their wants.

Objective 2: Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use

This was achieved with the exception that, arguably, the interfaces failed to highlight clearly enough which items were shared among the team. Although this subtle short-

coming did not affect the usability within a region, it may have arguably reduced the potential for team-wide negotiation of strategies (see Objectives 6 and 8 below)

Objective 3: Allow a single teacher to facilitate and debrief the activity, using an ecology of displays

Similar to 4Decades, the teacher's orchestration load was manageable. The simple handheld interface allowed the teacher to control the pace of the simulation, according to the available time. Automatic locking of the tablets between the rounds was found useful, especially throughout session 1 where the teacher took some time after every round to lead the class through reflection and comparison.

Objective 4: Encourage participants to reflect on assumptions through decision making and discussion

The activity provided groups with relevant, individual questions to reflect on, such as, *would you be likely spend £150 on a vacuum cleaner if you were a student living on a loan; how often do you think you would go out to parties?*

Objective 5: Allow participants to refresh from exhausting lectures by moving and talking

All students in both sessions were engaged, focused and energised throughout the simulation. There was much laughter and cheering in all the groups. The overall level of equitable participation was described by a teacher as remarkable, in contrast to previous experience with group activities at the same age.

Objective 6: Encourage negotiation by letting participants follow different goals but requiring the consent of teammates

Purchase decisions were discussed exclusively *within* regions, rather than between regions, as observed consistently in both classroom sessions. This observation resulted from applying the Three Bubbles rule of thumb while watching the group as a whole: Although regions shared and the class compared, there was a clear lack of team-wide mutual awareness. Within the strict definition of teams and regions (in Three Bubbles

terminology), *consent of teammates* was not sought at any point. Rather, *consent of region-mates* was the factor that counted.

This finding was a surprise, since the design had anticipated that regions would negotiate the purchasing of shared items, such as TV, toaster, etc. To explain the finding, it is plausible that the incentives (i.e., team-shared ownership of some items) were not highlighted clearly enough to be salient for participants during decision-making. Although the shared items were explicitly pointed out in the ambient display, the tablet interface did not distinguish them from other items. Since the decision-making was mainly focused around the tablet, arguably this is where the shared items should be highlighted explicitly. Future versions of the interface should address this possibility.

To further increase the incentives to negotiate, a teacher suggested to include an option for regions to split the cost of shared items, such as a TV. This suggestion is worth examining in more detail, since it raises a number of questions that are relevant to designing CLOVER applications more generally. Specifically, it calls for clarification regarding the necessity, feasibility, benefits and cost of Multi-Plane Groupwork (MPG) in different settings, as discussed below.

1. **Necessity:** First it must be clarified whether a pedagogical need for team-wide negotiation really exists. In both classroom sessions the collaborating groups were 5-7 students in size. It may well be argued that this size is appropriate for the activity, given its focus on *personal* (rather than global) spending decisions. In contrast, 4Decades had emphasised the challenge of building global consensus among leaders from diverse organisations. In this context the goal was *getting everyone to talk with everyone else*, whereas UniPad's explicit goal was merely *getting everyone to talk* which was clearly achieved. Therefore, arguably there was no need for MPG in the case of UniPad. However, for the sake of discussion it is worth exploring the idea further.
2. **Feasibility:** Suppose there was a need to encourage MPG in an application the size of UniPad. This would imply several challenges for a group to overcome, as the case of 4Decade has shown, in order to achieve effective and inclusive MPG. These can be thought of as four steps.
 - a) First the regions must learn how to control the simulation from their local, regional perspective before it makes sense to talk strategies.
 - b) Second, regions must then realise the incentives for MPG.

- c) Third, regions must buy into these incentives.
 - d) Fourth and finally, they have to figure out effective and inclusive spatial formations within the given constraints of the room. The case of 4Decades showed that these four steps were not always trivial to achieve. The fact that MPG typically did not start until the second match indicates that steps 1 and 2 require some time or indeed several matches. UniPad is relatively short (15-25 minutes) and the workshop time constraints have so far allowed only a single match. Step 3 pertains to the *amount* of incentive. This is what the above suggestion from the teacher aimed to increase. Step 4 can be risky, as one of the 4Decades sessions showed, where the group ended up with some problematic formations (session 3, see Chapter 7). Whereas the observed spatial formation in UniPad's two classroom sessions were found to work well regarding engagement and inclusion, it is difficult to predict what would have happened if the regions had decided to move their tablets closer together. Considering the large classroom sizes of 23 and 26 students, it is worth exploring whether MPG is potentially feasible in general using the minimal setup of only 4 tablets in total. Because in order to talk across the table (like in 4Decades), teams of 13 students would have to somehow squeeze around two tablets for extended periods of time. This points to another challenge for future work to explore, especially in real classroom settings where given table formations may not always be ideal. Using the above four criteria for feasibility, a systematic study might investigate, e.g., what is the minimum number of tablets that can support effective and inclusive MPG in a class of N participants during a time T , for a variety of relevant discussion tasks.
3. **Benefits:** In Chapter 7, Section 7.18.0.1 I argued for potential benefits of MPG in avoiding regional disengagement by providing peer social control mechanisms. This feature was not needed in the two UniPad sessions, since all regions were continuously engaged with the task.
 4. **Cost:** Furthermore, it can be argued that MPG may potentially distract regions from their tablet-based task. This may not always be desirable, especially when the tablet task is itself challenging. Unlike 4Decades (where the essential task on the tablet was extremely minimalistic and somewhat repetitive) UniPad's tablet interface constantly provided new material, as each round involved a different set of purchase options. It is fair to say that each round presented new challenges that a region had to first debate among themselves - from their own point of view -

before it made sense to take any potential interests of other regions into account. Because the rounds allowed less than 2 minutes for discussing several new purchase options, conceivably the group were too pressed for time to think about other regions' intentions during the rounds.

To summarise the above considerations, the two classroom sessions found no instances of strategy negotiation between any two regions, as negotiation happened exclusively *within* the regions. Arguably, this did not affect the activity's essential pedagogical goal, which raised general questions regarding the need, feasibility, benefit and cost of MPG in different cases. The above questions and preliminary answers (based on comparing findings from UniPad and 4Decades) are relevant to future work in three ways: (1) by helping designers decide when to use MPG and when not; (2) by explaining groups' appropriation or non-appropriation of MPG in four steps: regional mastery, realisation of incentives, buying into incentives and spatial formation; finally (3) by suggesting starting points for future studies exploring viable spatial formations and incentives for MPG in general terms.

Objective 7: Support engagement and constrain collaboration effectively by exploiting social mechanisms, such as team play

The regions' behaviour and discussions suggested that they all identified with their characters in the simulation. Shared identification was indicated as individuals argued from a "we" perspective, e.g. "*We don't want the telly*" - "*Yeah we do!*"

Objective 8: Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data

The ambient display was used by students and the teacher in both sessions. Comparing each other's decisions allowed the regions to be mutually aware of each other's *actions*. However, since the ambient display was used mostly in hindsight, i.e. after the rounds, there was virtually no indication of mutual awareness of *intentions* between regions. The distinction between actions and intentions comes from Yuill and Rogers (2012) and is relevant in this regard.

As far as the field observations revealed, attention to other regions' decisions was paid only in hindsight (after a round), rather than during the rounds. This observation was similar to the early rounds of 4Decades, where regions also worked in isolation initially.

Whereas team negotiation started in 4Decades at some point during a session (typically in the second match), this did not happen in the UniPad classroom sessions (where only one match was played).

Objective 9: Allow all participants to contribute equitably to the activity and ensure that no-one is excluded

Inclusion and equitable contribution were found to be remarkably high in both sessions. This finding was consistent across two different groups, room conditions, teachers and facilitation styles. The only observed (minor) instance of spatial exclusion was described as a “J-Formation”. One student at the remote end of the “J” (group of 7) ended up being far away from the tablet that was shared mostly near the other end. The issue was considered minor, since the student nevertheless engaged verbally and seemed neither upset nor distracted.

Similar to the Corner-L observed in 4Decades (session 3) there was no obvious indication that the group was aware of the fact that one member was at a disadvantage. By contrast, the more symmetrical groups showed high awareness of peers’ spatial needs. This was particularly evident in session 1 where the regions sat in circles and frequently took turns holding the tablet.

9.6.8.2. Sustainability

The chosen approach stressed that the teachers should be able to set up and use the simulation without the researcher’s technical assistance. This allowed them to independently use their own testing methods and to familiarise themselves with the logistics of using the application in the wild. With regard to hardware, this included charging, transporting and handing out the tablets. Software-wise, they had to navigate to a specific URL and set the browser into full-screen mode for all the tablets and the projector. Under ideal conditions, these procedures were simple enough for teachers to perform on their own, in order to run the simulation under technically ideal conditions. However, internet access was identified as a potential bottleneck in the wild, raising further challenges, as outlined below.

- The IT infrastructure in a school may require new devices (e.g. tablets) to be registered before they can connect to the school’s own WIFI network.

- Bringing own tablets (like in the case of MyBnk) may thus make a teacher dependent on technical staff whose availability may not always be guaranteed.
- Some smartphones allow setting up a temporary WIFI hotspot, allowing other devices to connect to the internet. In some places where WIFI is not available per default, this method (known as wireless tethering) can be viable in some cases, as the test in the community centre showed. This possibility was enabled by the application's low use of data resources, as opposed to 'media-rich' applications.
- However, the tethering approach has its limitations, since mobile network coverage may not be available in all classrooms. Moreover, the use of tethering may be subject to restrictions by smartphone contracts and school IT regulations;

Limitations such as the above may crop up surprisingly in situ. Overcoming these obstacles may in some cases require more know-how and technical resources than can realistically be expected from "the average teacher". For instance, even the MyBnk teachers (who were relatively tech-savvy) may not generally know how to use tethering, let alone find out *ad hoc* whether it is allowed by their personal contract or the venue's policies. To ensure that a particular school (or room) will provide appropriate internet access for the simulation, one would have to actually go there in advance to try things out. Within the constraints of how MyBnk operates, this was not a generally feasible option. Therefore it would be understandable if some of the teachers preferred to use a "safe" (e.g. paper based) option rather than taking a risk of unexpected technical delays.

9.6.8.3. Use of CLOVER

Out of the 8 CLOVER tools, 7 were validated by UniPad, i.e., all except the Benefits of Rule-Changing. The latter feature was not needed in the case of UniPad, since the given time constraints of the workshop allowed only one single match. Apart from this limitation, UniPad succeeded in validating all the CLOVER tools in a classroom context (Table 9.4).

The design was informed by Triple-i, which was found conducive, partly because it was compatible with MyBnk's usual approach to pedagogical innovation. The 9 Gameplay Objectives were essentially satisfied in two classroom sessions which applied the sequence of Lecture Sim Discussion. Three Bubbles provided the overall activity structure, by associating regions with devices, teams with strategies and the whole class with ambient information. Multi-Plane Groupwork (MPG) was applied in the design but not appro-

| N | Name | Usability with teacher | Usefulness in design/evaluation |
|---|---------------------------|--|--|
| 1 | Model Sim Game | “Model” was found too abstract initially. Some ambiguity between “Simulation” and “Game” | “Plugin scenario” informed design plans for adapting the content to different learner groups |
| 2 | Lecture Sim Discussion | (no problem) | The course organisers agreed with the pedagogical rationale. The simulation was used towards the end of workshops after lectures |
| 3 | Triple-i Hierarchy | (no problem) | This was how they worked usually anyway |
| 4 | Gameplay Objectives | Teachers did not need to use this tool in great detail | very useful in design and evaluation |
| 5 | Three Bubbles | Teachers preferred the concrete terms “character” and “flat”, rather than “region” and “team”. | Informed the basic structure of regions, teams and class. Useful for realtime evaluation and post analysis |
| 6 | Multi-Plane Groupwork | Occasional ambiguity, as teachers said “team” to mean teams or regions | Informed the idea to update ambient information instantly. Not appropriated by participants |
| 7 | Spatial Formations | (no problem) | Useful for discussing findings and thinking about future designs with more tablets or MPG |
| 8 | Benefits of Rule-Changing | (not used) | (not used) |

Table 9.4.: Overview of findings regarding the use of CLOVER in the design and evaluation of UniPad

priated by participants. A salient finding in this regard was that the activity remained functional despite being “downgraded” by participants to Single-Plane Groupwork.

Moreover, Three Bubbles was found extremely useful for evaluation. During field observation of whole-class activity, the model’s rule of thumb helped the researcher (myself) remain focused on the critical factors: *“Regions should share, teams should be aware, and the class should compare.”* This focus allowed the researcher to detect subtle aspects of group awareness (e.g. actions vs intentions) and spatial inclusion (i.e. the J-formation) in the midst of intense, bustling interaction: groups shouting, cheering, arguing, joking, laughing. Having a concise set of relevant indicators to watch

out for was effective for avoiding getting distracted. Although this approach did not produce (nor rely on) a detailed description of the session, it was nevertheless found suitable for verifying the activity's overall feasibility. For this purpose, the method may be recommended as a lean alternative to tedious, bottom-up video analysis in future evaluations of CLOVER-based (or CLOVER-like) activities.

9.6.9. Conclusions from the UniPad study

The UniPad activity provided groups with prompts regarding important question to reflect on, such as, *do you think you would spend £150 on an automatic vacuum cleaner if you were a student living on a loan? How frequently would you go out? If you had to choose between the Halloween ball and paying your the rent on time, how would you choose?* Rather than asking these questions on a paper sheet, the interactive simulation provided a framework for making these decisions in a social context with peers in the same situation. Keeping an eye on their bank balance round after round helped students engage with the idea that spending decisions have consequences, practice long-term thinking and pay for needs before prioritising wants. Shareable tablets were found suitable for negotiating these decisions in groups of up to 7. Realtime ambient information helped the teacher engage usefully in this process and supported groups' mutual awareness and joint reflection.

The study has achieved its academic goal of demonstrating an application that can be used sustainably in the wild (Kreitmayer et al. 2013b). Moreover, the study provided support for multiple aspects of CLOVER. By following the CLOVER Gameplay Objectives and many of the recommendations, the design quickly arrived at a usable application that replicated the essential benefits of 4Decades in a different workshop context. Although the given workshop context was too time-constrained to conduct collaborative rule-changing with students, the notion of “pluggable scenarios” was nevertheless welcomed by the course organisers, as it allowed for adapting the activity to different target groups.

Moreover, the inherently difficult challenge of visually assessing large-group collaboration was notably facilitated by the Three Bubbles model. Due to its manageable set of relevant indicators, a researcher was able to perform this assessment efficiently in real time. This eliminated the need for time-consuming video analysis and furthermore raised the question whether teachers could use the same method for monitoring a group during orchestration.

Future work should explore the boundaries of realtime visual assessment using Three Bubbles, including more challenging settings where groups walk around. Moreover, future evaluations of activities like UniPad might identify a greater range of spatial formations that support or threaten equal participation, thus extending CLOVER in useful ways. Finally, different workshop contexts with more time available might benefit from collaborative rule-changing. CLOVER's recommendations in this regard offer potential to inform new design and pedagogical strategies including, but not limited to, future versions of UniPad.

Furthermore, evaluating Gameplay Objectives 6 and 8, produced new insights regarding when to choose Multi-Plane Groupwork (MPG) over Single-Plane Groupwork. Relevant factors to consider include (1) necessity: *do the pedagogical goals require MPG?*; (2) feasibility: *is there enough time for MPG? Can the activity provide appropriate incentives for regions to negotiate? Can the room conditions support spatial formations that are appropriate for MPG?*; (3) benefits: *what do groups gain from MPG in a particular case? E.g. is building broad consensus relevant to the topic? Are regions at risk of disengaging without peer control?* and (4) cost: *e.g. might MPG distract from a regional task that requires focus?* These factors should be considered in future work to explore feasible spatial formations for MPG in classrooms with various furniture constraints. Although UniPad may remain as a Single-Plane Groupwork activity for the reasons discussed above, other applications may benefit from MPG.

With regard to sustainability, the setup was found easy enough for non-technical experts to use, given appropriate internet access. UniPad's '*media-frugal*' design (i.e., using only some text, numbers and few images, rather than large video or audio data) was found useful in this context, as it did not require the connection to be fast. These factors make applications like UniPad potentially scalable and sustainable in many classrooms around the globe (given that a projector, tablets and internet are available, which is increasingly common). Therefore, the web-based approach taken in this study was generally supported as a viable option in settings with a reliable IT infrastructure, highlighting potential to design more applications like UniPad that can easily be used in tablet-enabled schools.

In the specific case of "travelling teachers" that bring their own technology (e.g., MyBnk) the dependency on quick and easy internet availability in situ is clearly a limiting factor. Future versions of UniPad might therefore explore the option of using a portable WIFI router, as used successfully in 4Decades (and ComfyBirds, see below). Although this would effectively solve the dependency on in-situ internet access, this solution adds another device that a teacher must remember to bring, learn how to connect, activate, etc.

Moreover, the solution requires bringing a laptop with the server application installed locally - which in turn adds considerable maintenance overhead, weight and dependencies on specific software and hardware devices. Although these problems may be solved easily in high-budget contexts (such as 4Decades), charities like MyBnk may not have the resources to hire technicians or train their teachers to become IT experts.

9.7. The ComfyBirds study

This section reports on the second design study that was conducted to validate CLOVER. As mentioned earlier, the design started with an existing application - called *JokeCity* - that allowed pairs of children (around 7-9 years old, and particularly for children with poor reading comprehension) to discuss language ambiguity, using a mouse and a vertical screen. Although measurable learning gains had been found in children after using JokeCity in pairs, the amount of human guidance required for the task limited the application's practical feasibility at classroom scale. To make the task orchestrable in schools, the study aimed to transform *JokeCity* into a whole-class activity, allowing multiple pairs to work in-sync, whilst a teacher can control the overall pace and provide adequate support for the whole class as needed. To realise this transformation, iterative design and evaluation in the lab and a school setting were used, aiming to (a) devise an appropriate activity structure and (b) gauge its feasibility and sustainability. CLOVER tools were found useful in addressing both these goals, as detailed below.

9.7.1. Aim of this study

The primary aim of this study was to explore how far the CLOVER abstractions can be extended to support different types of activities. To address this aim, another whole-classroom activity was designed - based on a different kind of task - and evaluated with regard to feasibility and sustainability under real conditions.

9.7.2. Method and setting

The design of ComfyBirds was based on an existing learning tool called JokeCity (Yuill and Bradwell 1998). Empirical evaluations of JokeCity had been conducted by Nicola Yuill (ChaT Lab, University of Sussex) who took part in this study as a researcher and

teacher. The main data sources in this study were design meeting minutes, the actual sketches and prototype, observations from lab tests and interview with the teacher after these tests.

9.7.3. Overview of the design and analysis

Figure 9.11 provides an overview of how the ComfyBirds study is presented below. The following subsections will explain in adequate detail the following stages of the design:

1. the original riddles task, including the JokeCity desktop application, that defines the pedagogical requirements;
2. the Triple-i-based design rationale and high-level objectives;
3. (detailed lab-based design iterations are omitted due to their minor relevance to the aim of this chapter);
4. the ComfyBirds prototype at a robust stage ready for evaluation with children;
5. early evaluations with children in the lab and a school, testing feasibility and sustainability;
6. a CLOVER-based discussion of findings from the evaluation and potential for future design.

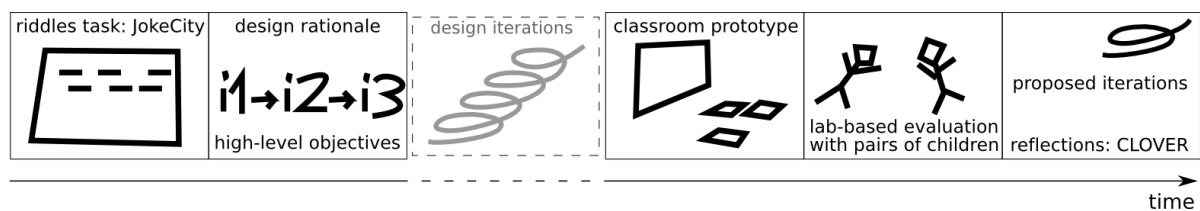


Figure 9.11.: Overview of the ComfyBirds study

9.7.4. The learning task: Joking riddles

The essential task was for learners to identify word ambiguities in joking riddles. For example, in the following riddle *“How do you make a sausage roll? Push it down a hill”*, the word “roll” is has two meanings, since it can refer to either a small bread product or a type of movement. Therefore, “roll” is the correct solution which learners were supposed to identify and articulate reasons why they think it is ambiguous.

9.7.4.1. Scaffolding in JokeCity

The JokeCity application was described in detail in Yuill (2007). It is essentially a single-user desktop application that was found to produce measurable increases of reading comprehension in children sharing the application in pairs (Yuill 2009).

Far from being a mere guessing game, JokeCity incorporated a pedagogical strategy of scaffolding children's thinking towards the correct solution. For this purpose, the graphical user interface guided the children through a series of stages for each riddle, as roughly outlined below and illustrated in Figure 9.12.

1. Initially, only the first part of the riddle is shown (e.g. *"How do you make puffed wheat?"*), allowing pairs to guess what the answer might be.
2. The second part of the riddle is shown (e.g. *"Chase it up a hill."*)
3. Participants are encouraged to click on the word that they think produces the ambiguity in the riddle .
4. A series of additional clues are provided on demand, such as (1) *"Think what makes you out of breath."*, (2) *"How do you make wheat puff?"*, (3) *"Click on a word in the question."*

9.7.4.2. Navigation in JokeCity

Besides guiding the user through each riddle, JokeCity furthermore put the user in control of general navigation, such as selecting different sets of riddles, proceeding to the next riddle after finding the correct solution, and quitting the application.

9.7.4.3. Disentangling and distributing JokeCity

One way to look at the overall design challenge is to untangle the three use cases: selection, scaffolding and navigation. This idea is graphically illustrated in Figure 9.13. It involves (1) constraining children's focus to the essential challenge of selecting the ambiguous word; (2) putting the teacher in control of the overall navigation using a digital (handheld) interface; and (3) letting the teacher perform all the required scaffolding non-digitally. Although the figure makes it look simple, the transition from desktop to classroom required a fundamental redesign.

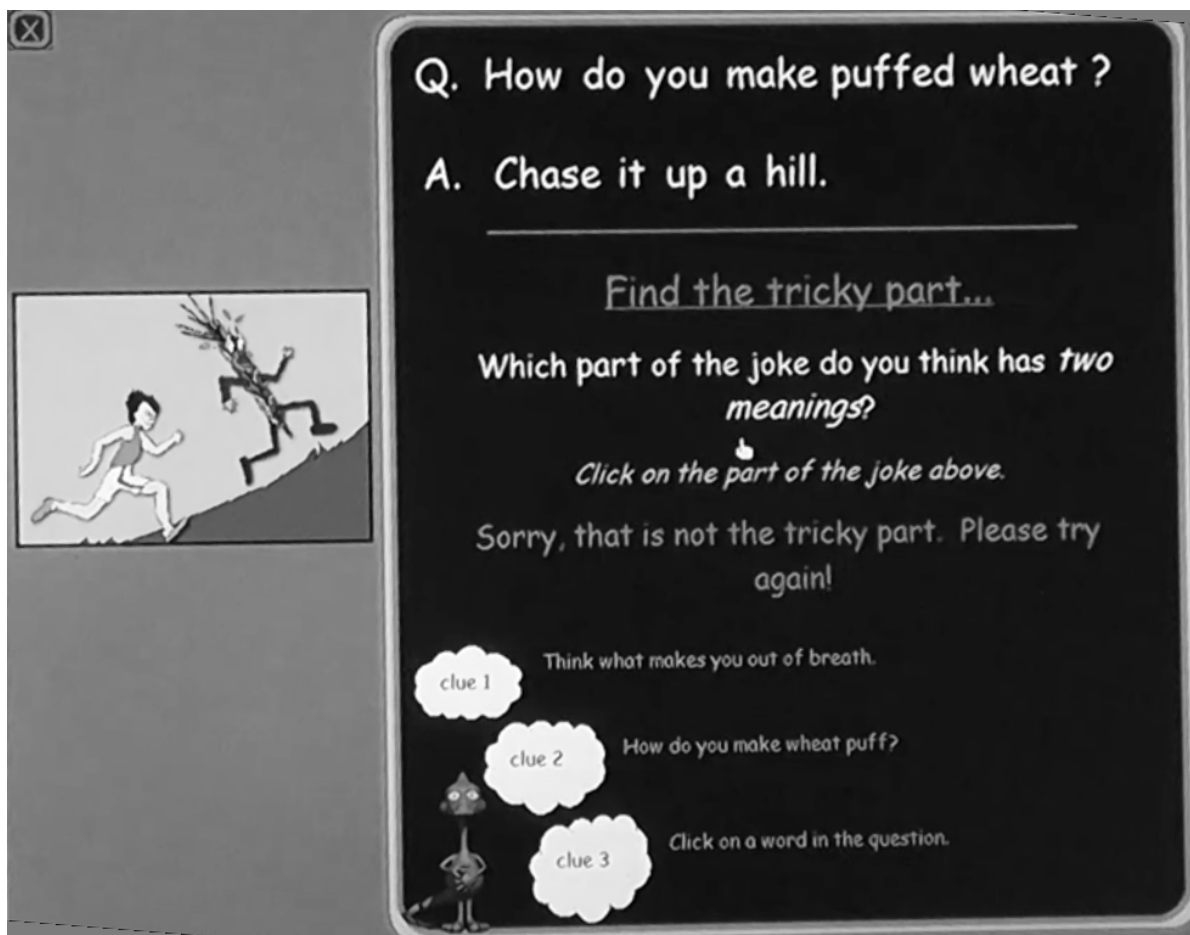


Figure 9.12.: Screenshot from JokeCity. The interface comprises (from top to bottom): The riddle, declaration of the task goal (i.e., “Find the tricky part...”), instructions (e.g., “Click on the part of the joke above”), feedback on correctness (e.g. “Sorry, that is not the tricky part. Please try again!”), and a little green figure that reveals up to three clues when clicked by the user.

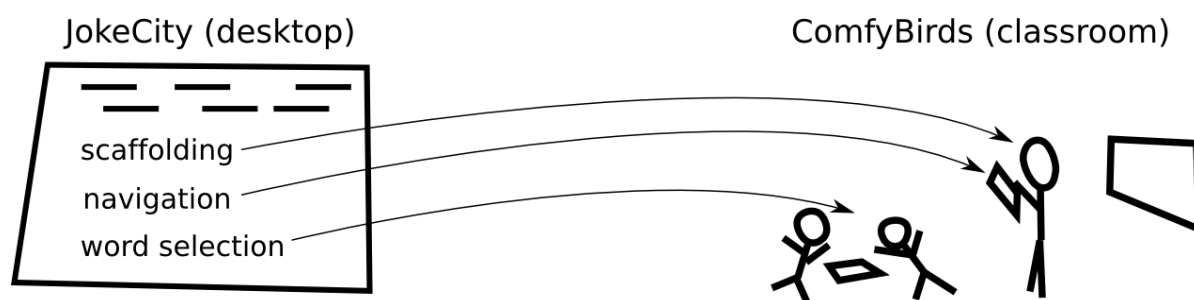


Figure 9.13.: Untangling the JokeCity application to a distributed classroom setup

9.7.5. CLOVER-based design rationale and objectives

The overall approach to design was outlined by Triple-i. As with the previous studies, this approach involved starting with the concrete instructional need (i1) before considering interaction structures (i2) and interfaces (i3) to address this need.

9.7.5.1. Use of CLOVER in the design

To devise appropriate i2 and i3 solutions for sustainable classroom use, several CLOVER tools and recommendations were used in the iterative design process.

- Gameplay Objectives informed the overall design goals;
- Several of the CLOVER recommendations guided concrete decisions;
- Three Bubbles, Spatial formations and Multi-Plane Groupwork were used in (a) inspiring the initial activity structure, (b) helping researchers and the teacher make sense of group observations and (c) indicating potential for design iteration.

Due to the nature of the task not being a simulation, several simulation-specific tools were not considered applicable, particularly *Model Sim Game*, *Lecture Sim Discussion* and *Benefits of Rule-Changing*. Moreover, some of the Gameplay Objectives were excluded for the same reason.

9.7.5.2. Design Objectives

The following aims were formulated to define the essential instructional need and some key requirements for sustainable use in classrooms.

1. to allow pairs of children to discuss and select the ambiguous words in joking riddles;
2. to put the teacher in control of navigating the activity, using a portable dashboard;
3. to let the teacher provide realtime scaffolding as appropriate (i.e., timing, amount, kind) according to his or her judgement of the group's situated needs;
4. to design appropriate visual metaphors and cues supporting (a) decision-making in pairs, (b) comparing and discussing with the whole class and (c) switching between pair and class level;

5. to implement the application using available hardware and web technology, allowing it to be easily portable, usable and maintainable.

To address these needs, some of the generic Gameplay Objectives were reused (Table 9.5), excluding the ones that were specific to simulations and team play.

| N | Objective | Description |
|---|--|---|
| 2 | instantly usable interfaces for 20-30 | Design an intuitive multi-user interface for 20-30 participants that takes minimal effort to learn and use |
| 3 | one teacher | Allow a single teacher to facilitate and debrief the activity, using an ecology of displays |
| 4 | reflect on assumptions via decisions | Encourage participants to reflect on assumptions through decision making and discussion |
| 8 | ambient visualisations for awareness | Encourage mutual awareness and extended dialogues among peers using real-time visualisations of simulation data |
| 9 | inclusion and equitable accessibility | Allow all participants to contribute equitably to the activity and ensure that no-one is excluded due to, e.g., physical factors, skills or emergent group dynamics |

Table 9.5.: Relevant subset of the the 9 Gameplay Objectives used in ComfyBirds

9.7.5.3. Ensuring orchestrability

To make the task orchestrable in large classroom settings, it was necessary to distribute information and control between teacher and students in ways that ensure mutual awareness, linearity and cross-plane integration. However, rather than copying the structure from 4Decades (which had been the essential strategy with UniPad), here the intention was for the design to stay close to the requirements of the original task.

Distributing interaction: 10 stages and 2 planes

While the learning goal required that children discuss in pairs, the teacher's feedback had to be collective to the whole class, rather than each pair individually. To reconcile these two needs, the activity was divided into separate phases that take place at pair

and whole-class level. Figure 9.6 lists these stages by name, description and plane. Two planes were used: class and region, whereby “region” refers to a pair sharing a device. 10 stages were identified as useful steps for guiding participants through a riddle. This list was initially only half as long and expanded as new stages were added over the course of several sketches, prototypes and lab tests. For simplicity, the final version is represented here.

| N | Stage | Description | Plane |
|----|-----------------------|--|---------------|
| 1 | Reveal question | The <i>question</i> part of the riddle is shown | class |
| 2 | Reveal answer | The <i>answer</i> part of the riddle is shown | class |
| 3 | First choice | Regions are allowed to select a word | region |
| 4 | Reveal first choices | All the selections are shown | class |
| 5 | Second choice | Regions are allowed change their selection if they want | region |
| 6 | Reveal second choices | All the selections are shown | class |
| 7 | Reveal correct | The correct word is revealed | class |
| 8 | Picture literal | An illustration is shown to explain one of the word’s meanings | class |
| 9 | Picture figurative | An illustration is shown to explain the other meaning | class |
| 10 | Razzle dazzle | A volunteer is encouraged to perform the joke in gestures | class |

Table 9.6.: One round in ComfyBirds comprises 10 stages, including 2 *decision stages* (bold).

Two of these 10 stages (stage 3 and stage 5) were designed to allow regions to make decisions, i.e., selecting which word they think is ambiguous. These are also referred to as *decision stages*. The whole cycle of 10 stages (that pertains to one riddle) is called a *round*. By comparison, rounds in 4Decades and UniPad comprised only two stages: one decision stage (aka “during the round”) and one reflection stage (aka “between the rounds”). Arguably, talking in terms of *stages* per round is a more precise and extendable terminology than saying “during / between the rounds”, which was really just a special case that happened to work for these simulations due to their simple time structure. By contrast, ComfyBirds’ time structure is far more sophisticated in this regard. Accordingly, a somewhat more sophisticated interface (compared to UniPad’s one-button interface) was required for allowing the teacher to navigate the activity. This

interface is described further below in Section 9.7.6.1. Similar to 4Decades and UniPad, a “match” in ComfyBirds comprised several rounds, i.e. in this case, a set of riddles.

Using CLOVER to validate initial design decisions

The distribution shown in Table 9.6 already defined a kind of wireframe (or high-level script) for the whole activity. Implementing each stage required fleshing out this wireframe structure with concrete, informed design decisions. CLOVER tools and recommendations were consulted for this purpose. Below I will highlight some notable questions raised by Three Bubbles and Multi-Plane Groupwork in this context.

Single-Plane or Multi-Plane Groupwork? Single.

The teacher’s rationale for having temporally separate phases for working in regions was that the given decision task was cognitively challenging for the target group. Therefore, the teacher suggested that during decision-making children should be focused in pairs, rather than getting distracted by the wider peer group. Therefore, it was decided to use Single-Plane Groupwork, rather than Multi-Plane.

Three Bubbles: Pairs and class - but no teams

Moreover, the activity structure so far did not involve a concept of teams - only regions and whole-class. In this respect, the ComfyBirds’ designed activity structure was partially, but not fully, consistent with the Three Bubbles model, insofar as it involved device-sharing in pairs and comparing as a whole class.

Three Bubbles: No long-term strategies

The concept of negotiating long-term strategies in teams (as suggested by Three Bubbles) did not apply in this case, due to the nature of the task. Particularly, each round presented a new riddle and the riddles were not connected by any kind of persistent state (unlike the simulations, where state was represented by multiple variables, such as money, temperature, possessions, etc.) In other words, decisions made in one round did not affect the situation in any subsequent rounds.

Spatial formations? Trivial.

The design of Spatial Formations in ComfyBirds was essentially straightforward. Front facing tables were required to allow equitable device sharing in pairs, whilst also making it easy for children to look at the teacher. Positive experience in the previous studies suggested the use of tablets (one per pair) and a single ambient display in the front.

Summarising from the above paragraphs, it is evident that the requirements of the case differed in some aspects from the context that CLOVER was originally intended for. As expected, some of the tools were more applicable than others. In any case, however, trying to apply the tools raised relevant questions. Answering these question enabled confident, well-reflected judgements as to why a particular approach was used or not used. In addition to the tools, a number of recommendations were followed as appropriate. These were summarised earlier in Table 9.1 near the beginning of this chapter.

9.7.6. Classroom prototype

Several months of design iterations, lab testing and consulting with other researchers and teachers led to the current version of ComfyBirds as described below. Rather than showing every detail of the application, I will explain just the most important aspects of gameplay, using relevant screenshots as illustration. Many details are not explained, such as how the teacher starts a new match, how the pairs initially pick a colour for their birds, etc.

9.7.6.1. Teacher dashboard

Figure 9.14 shows the teacher dashboard. It contains three buttons; one to restart the activity with a new set of riddles (“New Challenge”) and two for navigating through the script (“Back” and “Next”). A large “Next” button was suggested that is easy to hit without having to aim, allowing the teacher to keep an eye on the group.

The central part of the screen shows the *script*, with the current *stage* highlighted in orange. Based on test runs in the lab, the following information was found relevant for the teacher to see on the dashboard in real time.

- The name of the current set of riddles (In the figure near the top: e.g. “1. roll, 2. bowl, 3. stand”, named after the ambiguous words in each riddle);
- The current number of participating tablets (aka “players”);
- The current stage within the current round (e.g. “FIRST CHOICE” in the figure);
- A concise reminder of the teacher’s key action in each stage.

Other relevant bits of information were found easy for the teacher to infer from other information sources. For instance, while the teacher interface did not explicitly indicate

which round (i.e. which riddle) was currently played, this information was permanently available by looking at the ambient display. For the purpose of keeping the dashboard simple and clean with large text, it was decided not to include any redundant information in it, unless a real need was indicated during practical use.

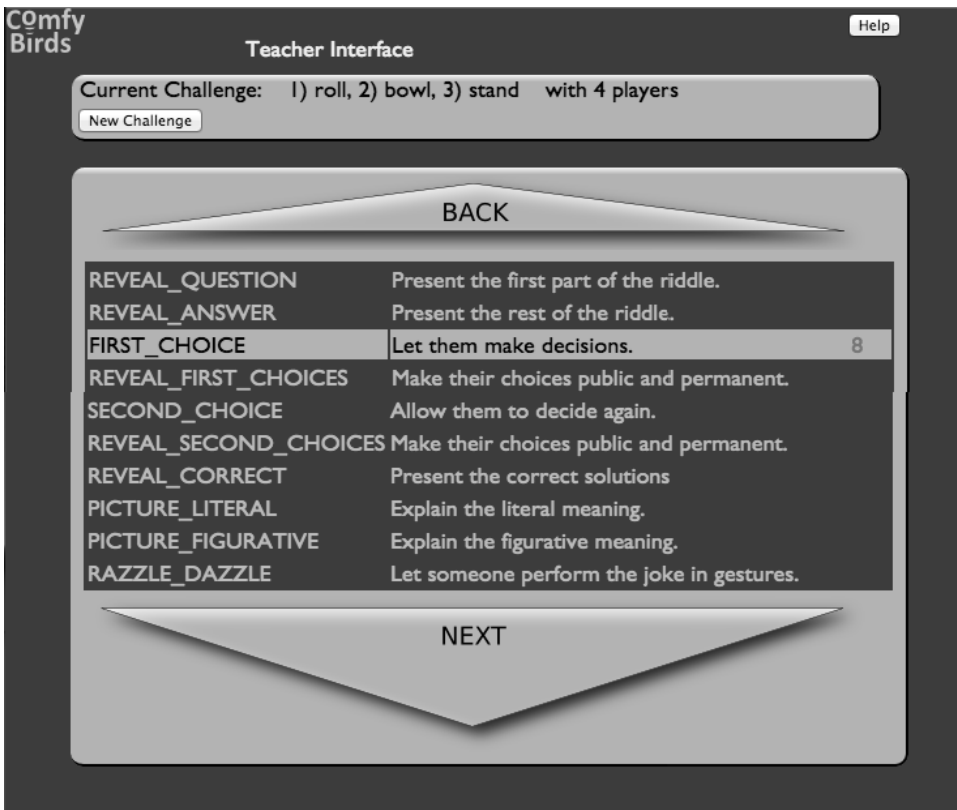


Figure 9.14.: The teacher dashboard: Large navigation buttons and a list of stages, highlighting the current stage

9.7.6.2. Region interface

Figure 9.15 shows a screenshot of the tablet interface. There is a longer story behind the name “comfy birds” and why they live on a sofa. Suffice it to say, for the purpose of this chapter, that the number of seats on the sofa corresponds to the total number of birds, which in turn is twice the number of joking riddles per match (in this case: 3 riddles times 2 decision rounds per riddle). After each decision stage, one of the birds flies off onto the ambient display. Therefore, the number of birds left on the sofa provides an indication of overall progress through the match.

The tablet interface allows a pair of children to select the word that they think is ambiguous, by tapping on it. The word is then highlighted (with a coloured underline,

see the figure) to indicate selection. The pairs can change the selection as often as they like, until the moment when the teacher navigates to the next stage - which renders the current selection final (like in 4Decades and Unipad and unlike the Desktop version, where clicking meant a final decision. The latter had caused occasional fights among the pairs about who gets to click). On the teacher's cue, the current bird flies up towards the selected word, while the screen fades to the "Look up" sign (Figure 9.16). The bird then appears on the ambient display (among the other pairs' birds, see Figure 9.17) and lands on a cloud that represents the selected word.

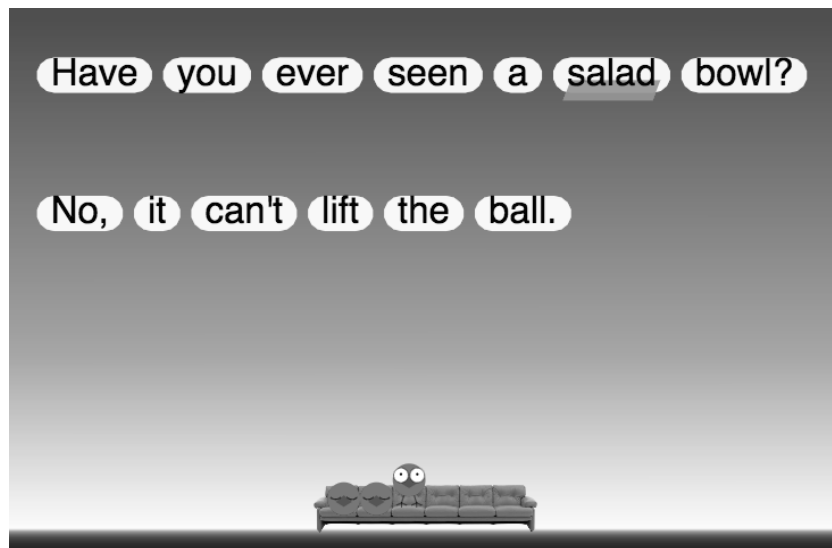


Figure 9.15.: The tablet interface with the word "salad" selected

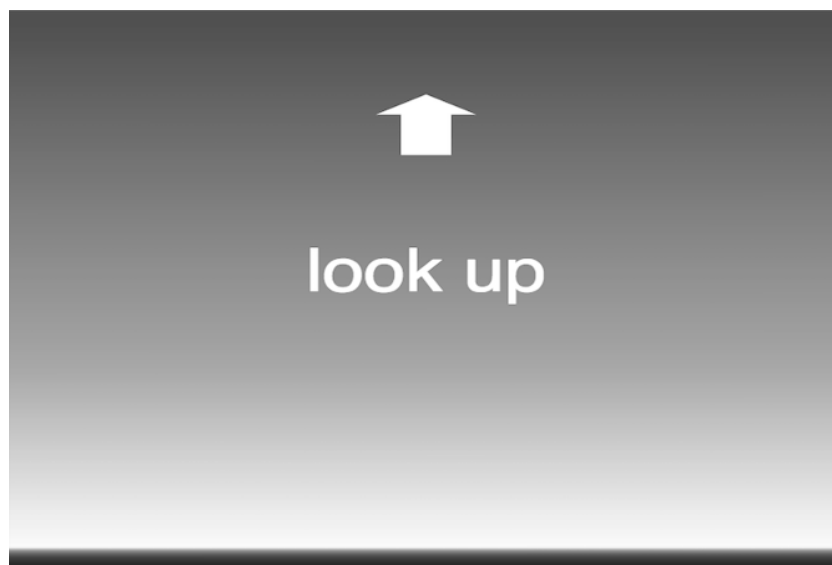


Figure 9.16.: The tablet interface directing attention to the wall display

9.7.6.3. Ambient display

Figure 9.17 shows a screenshot of the ambient display, with the joking riddle represented as clouds. As the teacher switches to the “Reveal choices” stage, all the birds simultaneously fly “off the tablets” and “onto the wall”. This transition is emphasised by (a) tweeting sounds, prompting the children to look up to the ambient display and (b) The “look up” signal on the tablet, as shown above in Figure 9.16. Each bird then lands on whatever word the pair has chosen. Each pair has a different colour of birds, allowing them to identify “their” bird easily.

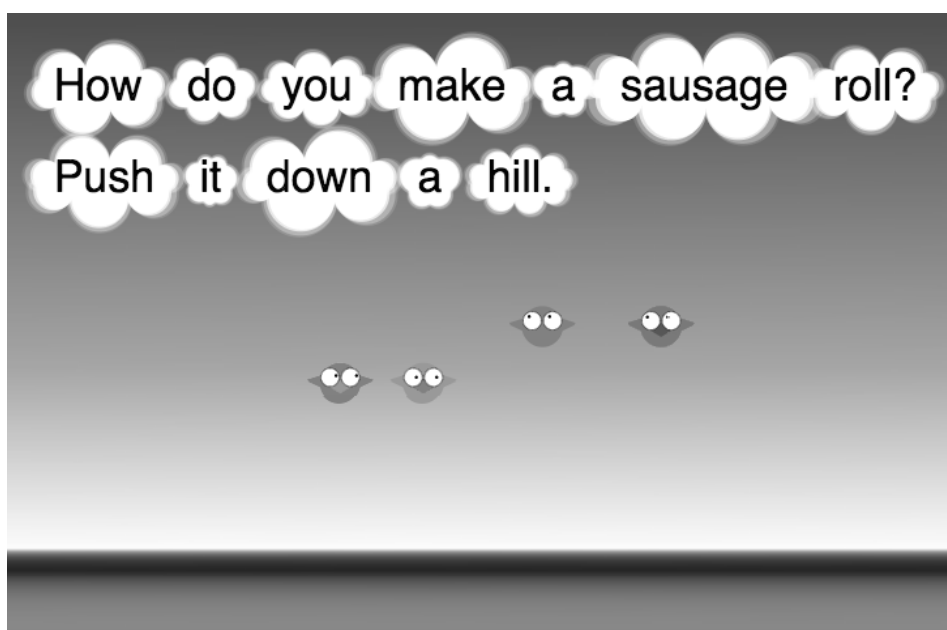


Figure 9.17.: The ambient display while transitioning from pair focus to whole-class focus

Figure 9.18 shows the ambient display in the stage where the “correct” word (i.e. the ambiguous word) is revealed to the whole class. It is visually highlighted by the corresponding cloud performing an animation. Any birds sitting on the “correct” cloud will also animate, by jumping up and down while flapping their wings. Later stages involve graphical illustrations to explain the two ambiguous meanings (Figure 9.19). As mentioned earlier, the activity involves 10 different stages for each riddle. Only some of them are explained here, to provide a rough overview of the essential procedure.

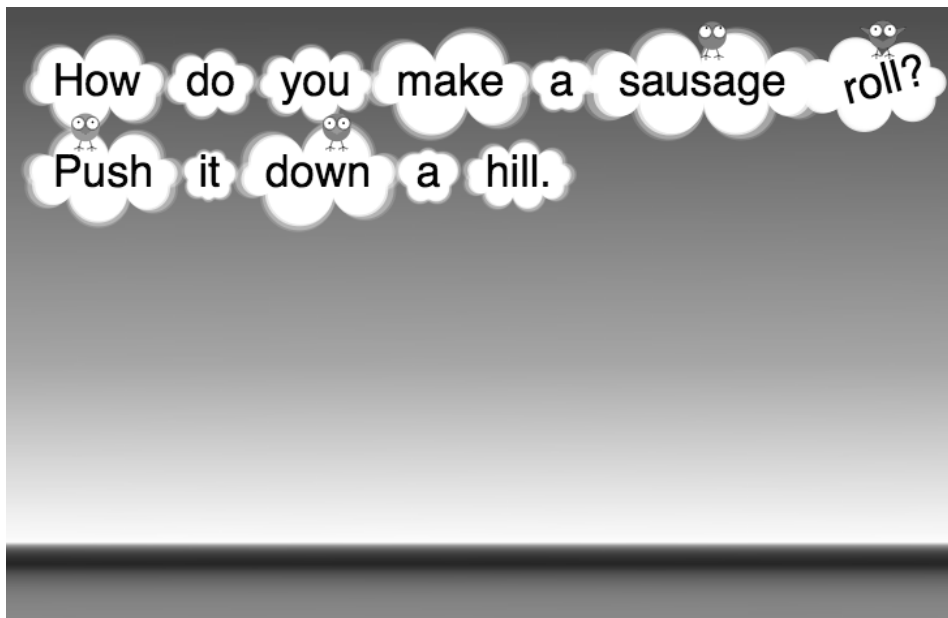


Figure 9.18.: As the correct word (“roll”) is revealed, its could animates and the green bird shows physical expressions of glee.

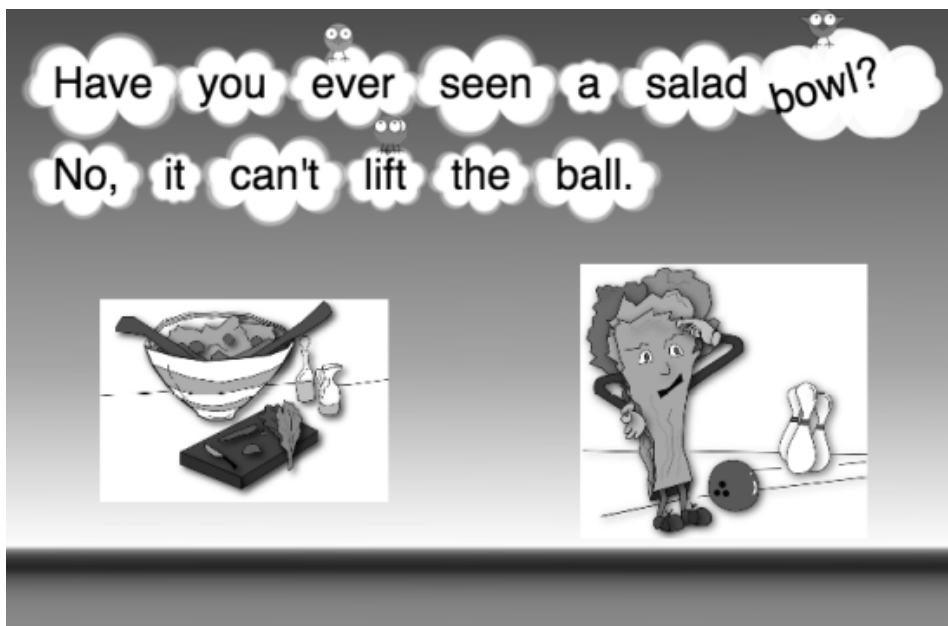


Figure 9.19.: Graphical illustrations explain the ambiguous meanings on the ambient display.

9.7.7. Evaluation

During the iterative design, most of the tests were conducted with other researchers in the role of children. This approach was sufficient for early prototyping, testing the technology and allowing the teacher to practice facilitation. Moreover, several teachers, other researchers and visitors to the lab tried the application in the lab and provided helpful feedback. This included a specialist literacy teacher who argued that (in terms of student-teacher interactions) the ComfyBirds structure exactly mirrored what a teacher would do *without* technology. Furthermore, teachers emphasised the specific benefit of letting pairs make selections without giving them the power to advance through the application. It was argued from experience with sharing desktop software that some children tend to fight about who presses the button that leads to a new page, especially when expecting a “well done” message or other kind of virtual reward. ComfyBirds’ teacher-controlled method of advancing was perceived as a useful feature in this regard.

At a point where ComfyBirds was considered generally robust enough, two evaluations were conducted with relevant target groups, using the prototype described above.

9.7.7.1. Evaluation with relevant target groups

The first evaluation was a lab test with two pairs of children (6-8 years old). The other was an in-the-wild study conducted at a special school with teenagers who were on the autistic spectrum (12-15 years old). Both sessions were discussed afterwards with the teacher in order to reflect on the design assumptions and how the design was appropriated by the various groups. The second study furthermore aimed to measure the activity’s effectiveness of increasing students’ reading comprehension, using a post test. The studies further showed that the application was generally usable, technically robust and easy to orchestrate in two different settings.

Evaluation in a special school setting

Evaluation at a special school was conducted independently by Nicola Yuill and colleagues at the University of Sussex. Although the designed intention was for pairs to share tablets, teaching staff at the special school insisted that each student should be given a tablet. A summary of the findings is represented below.

“12 students with ASC [autistic spectrum condition] had 3 30-min sessions of ComfyBirds in small groups (varying between 2-4 per group; a typical class size in this school is 6-8). Each student completed pre-tests and post-tests: a riddle understanding test (shown to correlate with reading comprehension) and a reading assessment with separate scores for reading accuracy and comprehension. Overall the group showed no significant increase in reading comprehension, but significant overall increase in understanding word compound riddles - though we did not have the resources to include a control group. All students contributed verbally and through the software in the sessions and most expressed regret when the sessions came to an end. Students had varying communication difficulties so participation was an important measure of success. Video analysis focused on students’ attention during the sessions: durations of attending to the tablet, their peers, the teacher or the wall (screen). There were instances of students comparing answers with each other and occasionally playfully altering others’ answers. All except one student took some part in articulating the multiple meanings of the text: this kind of talk has been shown to predict comprehension improvement in typically-developing children.” (Nicola Yuill, 2014)

With regard to sharing, it was found that, although each student had their own tablet, there was a surprising amount of interaction between students:

“[...] they did take notice of others’ screens, which is nice in itself, especially one pair who were girlfriend/ boyfriend -they even did a bit of joky switching their partner’s birds (especially nice in this group since nice teasing is not a very ASC behaviour).” (Nicola Yuill, 2014)

Several students asked repeatedly when they would be able to get ComfyBirds on the App Store. This was consistent with experience from other audiences, including lab visitors and conference audiences.

This in-the-wild evaluation corroborated lab-based findings regarding ComfyBirds’ overall feasibility and adaptability to different target groups. Moreover, the post test indicated its effectiveness as a learning tool. Finally, the activity’s flexibility and technical robustness indicated potential for sustained use in special schools and mainstream schools.

Lab session with young children

Whereas the above evaluation primarily showed that the application could be used in the wild, the lab session with typically-developing young children was especially revealing

with regard to validating CLOVER. Despite the small amount of data (only one session, only 2 pairs), it raised some challenging questions and indicated potential for extending CLOVER. Therefore, the following analysis focuses on the findings from this lab session, due to its relevance to the use of CLOVER as an analytical instrument.

During the entire session (3 rounds of joking riddles) the teacher was able to guide the group and provide scaffolding as needed. A surprising finding was that some of the children were remarkably curious to see what word the other pairs had selected, before the decisions went public. The findings are detailed below.

9.7.7.2. CLOVER-based analysis of group interaction

In the following, relevant findings from the lab session with the young children are described below using four CLOVER tools: (1) Three Bubbles, (2) Spatial formations, (3) Multi-Plane Groupwork, (4) Gameplay Objectives.

Three Bubbles

Similar to UniPad, the analysis started with the *rule of thumb*, verifying that, as expected, both pairs were able to share the tablets and compare decisions as a class (albeit a small class) using the ambient display.

Subsequent to the rule of thumb, the model's *5 arrows* short questionnaire was used to examine more closely the relations between observed behaviour, bubble structure and mechanisms for collaboration (Table 9.7).

Spatial formations

The children were seated in pairs at front facing tables, one table behind the other (Figure 9.20). The layout was mostly static, in the sense that the tables, chairs and tables were not moved. Occasionally, however, individual children ran over to the other table, apparently curious to see the choice on the other tablet. Moreover, one child from the back row chose to linger between the tables for a while, from where it was easier to see both tablets at once.

| N | Question | Answer |
|---|--|---|
| 1 | (How) did operating its own device provide each region with control over the interface? | Both pairs had an initial tendency to take turns, one controlling the tablet whilst the other was watching. Instances of equitable sharing were also observed. The tablets were not moved from the table. No other participants but the pair controlled each tablet. |
| 2 | (How) did sharing the devices <i>within each region-bubble</i> support mutual awareness of (a) actions and (b) intentions? | Pairs were engaged in discussing which word to select. |
| 3 | (How) did negotiating strategies <i>within each team-bubble</i> support mutual awareness of (a) actions and (b) intentions? | (strategies were not supported by the application) |
| 4 | (How) did comparing simulation data <i>within the whole class-bubble</i> support mutual awareness of (a) actions and (b) intentions? | The teacher encouraged and facilitated comparing between decisions. Alternative explanations for correct solutions were discussed, whilst keeping the discussion of incorrect solutions brief. In an email interview the teacher stated: <i>“It is usually not that profitable to talk for much time about a wrong guess so I tended to encourage discussion about correct choices but not about incorrect ones. There’s a nice example where [anonymised] makes a wrong choice [...] and I discourage him from discussing it (as it would be a dead-end).”</i> |
| 5 | (How) did interpreting simulation data <i>within the whole class-bubble</i> make background information available throughout the activity? | The ambient display supported awareness of <i>actions</i> (not intentions) by showing decisions in hindsight. The only way for a pair to be aware of another pair’s <i>intentions</i> during the decision stages was to physically look into the other pair’s tablet screen. |

Table 9.7.: The “5 arrows” questionnaire, applied to the ComfyBirds lab session with young children

Multi-Plane Groupwork

In order to determine whether or not Multi-Plane Groupwork took place, the important question to ask is *“Were there any relevant interactions between participants across*

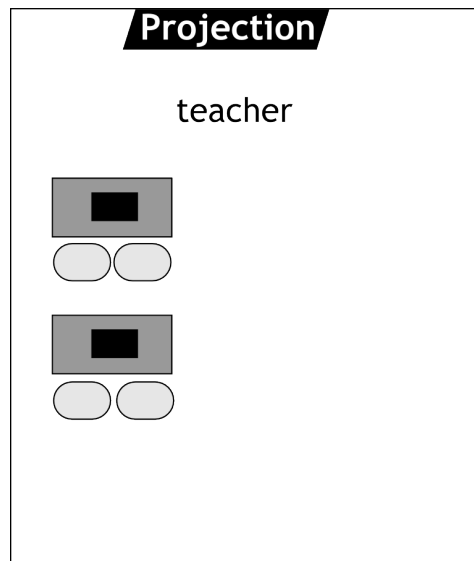


Figure 9.20.: Room layout in the lab test. Two pairs facing the teacher while sharing tablets

multiple planes at the same time?” In regard to this session the answer is yes, as children could not always resist the temptation to run over to the other pair to see what they were choosing. This happened repeatedly despite the teacher kindly asking the children to stay seated. Apparently there was much curiosity about the other pair’s intentions. Since the application did not provide for this need, children found their own way of satisfying their need for *awareness of others’ intentions*, by physically moving between tables.

Gameplay Objectives

Table 9.8 reflects on each of the Gameplay Objectives that were considered relevant to ComfyBirds. All of them were found to be satisfied. Moreover, the activity satisfied all of the specific design objectives formulated in Section 9.7.5.2.

9.7.8. Potential for design iterations

Overall the design was appropriated by the participants as expected - with the exception of children wanting to see the other pair’s tablet, which the design had not anticipated. Rather, a conscious design decision had been made in favour of Single-Plane Groupwork. The observation that the children essentially turned it into Multi-Plane Groupwork was surprising. It indicated a salient need (or at least a strong interest) of children to be

| N | Objective | Satisfied | Comment |
|---|--|-------------|--|
| 2 | instantly usable interfaces for 20-30 | as expected | The tablets were usable in pairs. Nothing should prevent the application from scaling up beyond the 2 pairs that were tested in the lab. |
| 3 | one teacher | as expected | The teacher was able to facilitate and debrief the activity, using the handheld interface and the ambient display |
| 4 | reflect on assumptions via decisions | as expected | The task triggered intense discussion about the word ambiguities |
| 8 | ambient visualisations for awareness | as expected | The ambient display enabled mutual awareness of actions (but not intentions), as decisions were shown in hindsight. |
| 9 | inclusion and equitable accessibility | as expected | All participants were able to contribute equitably. |

Table 9.8.: Findings according to the relevant Gameplay Objectives

aware of peers' intentions, thus raising the question as to whether the design should accommodate this need.

Solutions worth exploring might include, e.g., making one or both decision-stages “multi-plane” by design. This may or may not involve a team structure. Particularly, it does not require a full Three Bubbles structure. Three Bubbles and MPG are remarkably independent in this regard, as ComfyBirds demonstrates.

9.7.9. Overall sustainability

The technical implementation of ComfyBirds (as a web-based application) was the same as in UniPad. Therefore, the two applications could be regarded as identical in terms of preparation and maintenance requirements, given the same number of tablets. Four tablets were used in the reported study in a special school. Mainstream classroom settings would require about 10-15 tablets.

Using four tablets the teacher was able to conduct ComfyBirds independently in various settings, including lab-based tests and demonstration at a conference. Stable internet connection was guaranteed in these settings. Therefore, the preparation and maintenance

load consisted mainly in preparing a bookmark to the game URL on each tablet and ensuring that the tablets were charged. This effort was perceived as manageable and did not require any advanced IT skills. Plans for rolling out the application at a large scale in classrooms (using the web-based approach) were corroborated by the application's technical robustness and easy orchestrability success.

9.7.10. Reflecting on CLOVER

To conclude the analysis, it is fair to say that the lab session, although small, raised a number of relevant questions that illuminated some of the CLOVER tools from different perspectives. These different perspectives were helpful in discussing some of the conceptual boundaries, such as the possibility to use MPG and Three Bubbles separately from each other. Therefore, the approach of using an edge case for additional validation was supported.

9.7.11. Conclusions from the ComfyBirds study

The ComfyBirds study demonstrated how CLOVER was helpful in transforming a desktop-based task into a feasible and sustainable whole-class activity. A surprising finding was that children apparently desired more other-awareness than the design offered them. Particularly, the findings highlight the importance of distinguishing between awareness of *actions* and awareness of *intentions*. Potential was indicated for Multi-Plane Groupwork to satisfy a specific need for awareness of *intentions*. This raises further questions for design, such as how MPG could be effectively constrained such as to satisfy the need on the one hand, whilst on the other hand allowing regions (e.g. pairs) to focus on a task if needed.

ComfyBirds demonstrated an edge case that deliberately aimed to extend the original scope of the framework. The results arguably benefited CLOVER as well as ComfyBirds.

9.8. Discussion

The following discussion reflects on the findings and conclusions from both studies, UniPad and ComfyBirds.

9.8.1. Multi-Plane Groupwork tends to surprise

Table 9.9 contrasts the three case studies with regard to the relation between Single-Plane Groupwork (SPG) and Multi-Plane Groupwork (MPG). For each case study, the table shows how the design was intended, versus how it was appropriated by groups in the wild. For instance 4Decades sessions consistently began as SPG and then expanded to MPG.

| Case | 4Decades | UniPad | ComfyBirds |
|--------------|----------|--------|------------|
| Designed | MPG | MPG | SPG |
| Appropriated | SPG->MPG | SPG | MPG |

Table 9.9.: The three studies compared by use of SPG and MPG

To explain why groups appropriate a design in one way or the other, this research provides empirical support for the following conjecture:

1. Successful collaboration relies on a well-constrained configuration of social, behavioural and technological mechanisms, as proposed by Yuill & Rogers (2012): (a) availability of background information, (b) mutual awareness of actions and intentions, and (c) control over the interface;
2. If a design constrains these mechanisms too strictly (see pair isolation during decision-making in ComfyBirds), groups will seek ways to circumvent these constraints;
3. Conversely, if the designed constraints are too weak (see team incentives in UniPad) then they are likely to be ignored.

If this conjecture holds true (as supported by the empirical studies so far) then the implications for design are: (a) use frequent user testing and iterative design; and (b) use the Three Bubbles model to diagnose whether the system offers appropriate constraints across all relevant levels of group size and during all relevant stages of the activity.

9.8.2. Varieties of Three Bubbles

Figure 9.21 compares how Three Bubbles manifested in the 4Decades, UniPad and ComfyBirds. The figure proposes a graphical shorthand notation that might be used during future field evaluation, as it could be used very quickly using pen and paper (or on a whiteboard during design). For instance, the dotted line in the centre indicates that the team structure in UniPad was ambiguous, as regions rarely expressed identification with their overarching team. The missing middle arrow shows that no team-wide negotiation took place. The right column shows that ComfyBirds had no team structure at all. This notation might be extended, e.g., by labelling the arrows (making it easier for novice users of the model to remember what the arrows mean). Moreover, there is potential for the notation to distinguish between *actions* and *intentions* with regard to awareness - as this has been highlighted as a crucial factor, especially in ComfyBirds. Finally, since these diagrams are small, they could be part of larger visualisations, e.g., analysing sessions as individual time slices (such as rounds, stages or minutes) and aggregating these into a readable whole.

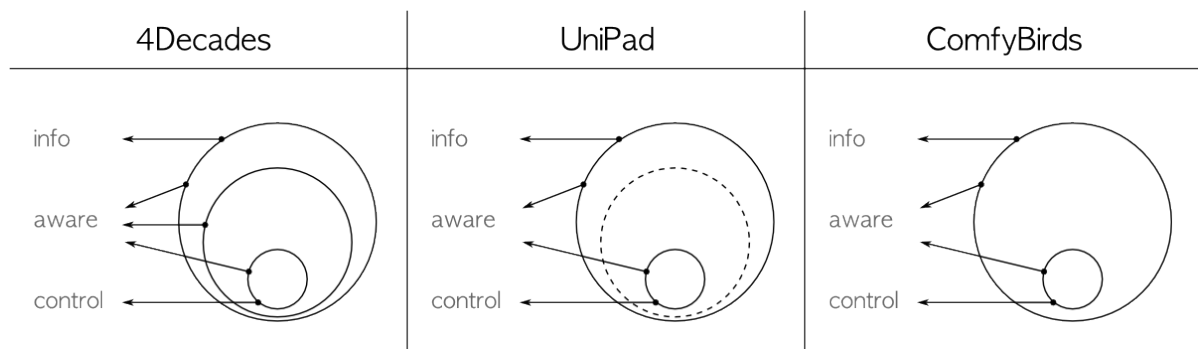


Figure 9.21.: Three Bubbles in 4Decades, UniPad and ComfyBirds

9.8.3. Potential for extending CLOVER

During the analysis it emerged that CLOVER (although useful and complete as is) offers potential for being extended. This includes, (1) more general recommendations regarding when to use MPG or SPG, respectively, (2) formulating the relations of *match*, *round* and *stage* a separate tool, (3) formulating *decision types* as a new tool, (4) augmenting some of the tools with specific questions to help constrain the analytical focus during analysis.

9.8.3.1. Guidelines for choosing between MPG or SPG

The UniPad study provided a deep analysis of the potential for using MPG and suggested four factor to consider: necessity, feasibility, benefits and cost. There appears to be potential for formulating these insights as a generic tool to be integrated with CLOVER.

9.8.3.2. A new tool for match, round and stage

Match, round and stage have emerged as useful concepts in all three studies. Together they offer a consistent set of concepts to describe (precise and concisely), the time axis of decision-oriented activity structures that emphasise Dillenbourg & Jermann's (2010) concept of linearity. Moreover, this new tool should contain a concept of "state", indicating whether the rounds are causally disconnected or there is potential for strategic planning across multiple rounds.

9.8.3.3. A taxonomy of suitable decision types

The three applications involved three types of decisions:

1. 4Decades: two integer values (mitigation and adaptation)
2. UniPad: a several-out-of-several selection (selecting multiple items from a given set of purchase options)
3. ComfyBirds: single selection from a set (selecting 1 out of N given words)

These three decision types are all different but they share in common some important features that make them usable in a CLOVER context, such as being small (sets or numbers), discrete and easy to articulate verbally (in discussion).

9.8.3.4. Adding specific questions to evaluation tools

During the studies it recurrently emerged that specific questions are useful to have at hand during analysis. For instance, the *5 arrows* questionnaire provided by the Three Bubbles was found helpful by asking the right questions. To get to useful results more efficiently, similar sets of questions might be added to other tools, particularly Spatial Formations and Multi-Plane Groupwork.

9.9. Conclusion

The two studies presented complementary approaches to testing CLOVER. While UniPad verified that the framework can guide new designs similar to 4Decades, ComfyBirds explored the framework’s adaptability to different cases. Benefits were found in CLOVER’s modular structure, i.e., the tools are compatible yet independent. This allowed designers to pick from the “tool box” only the tools that were relevant to each specific case.

With the exception of Benefits of Rule-Changing, all CLOVER tools were used in one or both studies, thus providing evidence for the framework’s relevance to a wider range of topics (from finance to language), discussion-oriented learning goals (global consensus, personal priorities and word meanings), classroom-like settings (high-resource and low-resource) and target groups (policy makers, children and teenagers). Figure 9.22 summarises this range, indicating potential for future work in four directions.

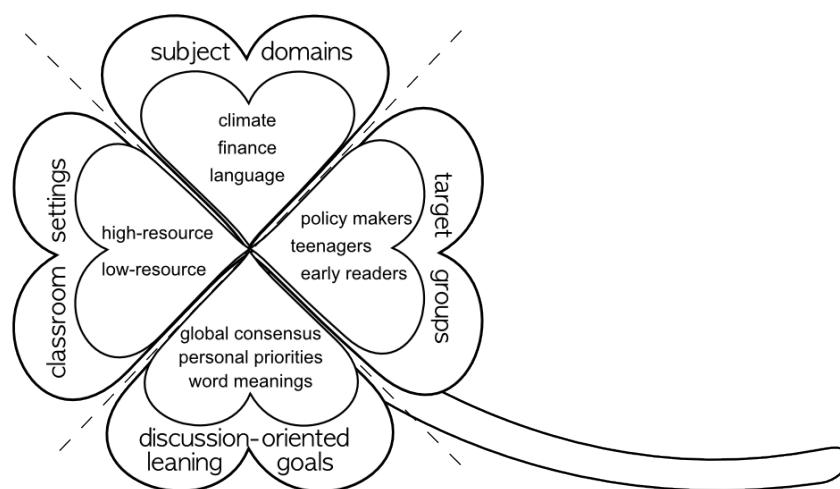


Figure 9.22.: Range of subject domains, target groups, learning goals and settings addressed by CLOVER so far

Within the range explored in this work, CLOVER proved helpful in identifying relevant needs, benefits, costs and constraints that informed design choices, leading to feasible and sustainable applications. Furthermore, some CLOVER tools were used to evaluate the activities’ feasibility, engagement and inclusivity. The Three Bubbles *rule of thumb* helped researchers analyse large-group engagement and inclusivity in realtime, suggesting that it could also be interesting for teachers to use during orchestration. More in-depth analysis was supported by the 5-arrows questionnaire that explained how in-situ appropriation was driven by the groups’ collaboration needs. Potential was found to extend CLOVER with new tools, particularly regarding time structures and decision types.

Appendix A.

4Decades rule-changing discussions

The following transcripts represent the discussions within two teams (A and B) while changing the rules in the 4Decades simulation. Represented are all four rule-changing instances in the second in-the-wild session (Matches 2-5). The context and procedures are detailed in Chapter 6.

For easy readability, the teacher's utterances are highlighted using bold text and line breaks. Indentation is used to approximately distinguish individual speakers.

Match 2. Team A chooses the winning condition

Highest income

Keep it as a planet, yeah

[reads the full screen out loud]:

Lowest temperature.

Base damage, don't you think you want the base damage?

Lowest base.

OK, you want base?

Temperature is probably easier.

I'm happy with temperature.

Lowest temperature?

It's easier.

Yeah, just to play with, especially in the beginning.

Yeah, exactly, to learn.

Having said this, actually it might be too easy because the lowest temperature is just mitigation, mitigation, mitigation - you know - five, five, five. Maybe base damage is more interesting.

Well, we can do it next round.

They'll probably choose in the next round.

Yeah because lowest temperature is basically what it all comes down to, isn't it - but not really because -
It's easy to play.

Teacher [to the team]: Have you chosen?

Not yet. O.K.

Temperature!

Teacher [to the class]: Let's have a look everybody at what Alpha Centauri's choice is for this match number 2.

So should we just maximise mitigation?

Yeah, but that's what it is. You do - yeah.

Maximise mitigation every round?

It's - yeah. It's max mitigation.

Match 3. Team B chooses the winning condition

I think we should go back to basics - money.

[many laugh]

That's what we're good at.

Teacher [to the team]: So that's your choice as a table?

So, ideal conditions are you have - global temperature stays below 2 degrees

yes

So we wanna have global temperature change under 2 degrees celsius

yes

And then maximise for money

Yeah.

Is that possible?

What are the options that we're getting here?

Is there only one or is it - so we have, erm -

[Reads out loud] The planet with the highest or lowest -

Only one option that we can choose from. So it's either the highest/lowest, income, emissions, temperature, baseline damage or residual damage.

Let's try base damage and see what that does.

Right, interesting.

Teacher: [to the team, addressing the participant who had made the suggestion earlier about 2 degrees] I heard that, that was a very interesting - you haven't got that choice. Did I hear you say you want 2 degrees and maximise income?

Several: yes

Teacher: So we could try and play that actually if that's what you want to do. That'll just mean that that screen will be wrong for match 3. So, sure, we could do that. Do you wanna do that?

Many: yes!

Can we do, like, 2 degrees or, like, 4 degrees?

Several: 2 degrees!

Teacher: You're gonna have to help me work out the winner. You'll see at the end of this - if we do it your way, which I agree is a very interesting way, we'll have to make out the winner ourselves.

[simultaneous discussion in the background, unintelligible]

Can we make 3, actually, our temperature ceiling?

Teacher: You can do it for 3 if you like. OK. Right. So just put anything in that iPad.

[Teacher starts announcing the winning condition to the whole group]

I like the idea of residual damage as well.

Teacher [to the class]: OK Betelgeuse has chosen their winning condition, which on the screen is written like this - ignore that, it's rubbish, right? So, we'll do it this way. They want to play a bit more sophisticated game, which is quite interesting. [anonymised - name of the participant that had made the suggestion], explain what the winning condition is.

Match 4. Team A chooses the scenario for mitigation and adaptation

Teacher: Now we're gonna give you some choices about the maths behind this game. [...]

OK, so we have the choice of being [reads] a techno optimist, business as usual and fossil fuel abundance.

techno optimist.

[several agree]

OK techno optimist

That means that we'll do lots of adaptation. Huge lots.

We'll do lots of mitigation, right?

Well, no, adaptation. Because -

Because we're techno optimist we think we'll be able to solve the problem.

Teacher [to the team]: So the options you've got there - the one we've played was business as usual. Techno optimist, you know, cold fusion tomorrow. There's tons of negative net present value [...]

[while the teacher is talking] That's more adaptation, right?

This is more *mitigation*.

Teacher: [...continued] or if there is a collapse in fossil fuel price then clearly mitigation costs just go right up. So, what world would you like to play?

I think techno optimist

[many agree]

Sounds good

Yeah I like that.

So the third gigatonne -

The third gigatonne you'll break even

Whereas normally you'd break even at the [unintelligible]

So if we, and that one we'll break even on the second.

So that means we can afford a lot more mitigation.

Several: Yeah / That's right / Exactly

Let's go for that one.

O.K., so we ain't gonna need to fix climate change, we can just adapt.

Well, erm, no.

You still do, but it's cheaper now to...

invest?

Several: into new technologies. / into mitigation.

than it was before.

Correct, correct. Mitigation is cheaper than adaptation.

Not cheaper necessarily, but cheaper than business as usual.

Several: Yeah / Mhm / Right.

So what's our strategy for this pass?

[The menu switches to adaptation selection]

There's an adaptation policy we have to pick, too.

Oh, OK.

What? Something else?

[reads out loud]: Every trillion invested in adaptation reduces the residual damage by -

[reads the thresholds etc]

Let's go for that one, let's just see what that changes.

It would be more interesting to go low. High is the sexy one but low is probably the more interesting one.

Let's go low then.

OK low.

Which means that we mitigate

Again we mitigate

Yeah we have to mitigate

Alright let's go for that.

Teacher: Done? OK

Match 5. Team B chooses the scenario for mitigation and adaptation

Techno optimist, business as usual and fossil fuel abundance.

I suppose that's the only one we haven't done.

Yeah let's try it.

No it doesn't change the cost.

So it just increases the cost of mitigation.

Right

[unintelligible]

Carbon emissions cost more to reduce

We haven't played fossil fuel

We haven't done that one, yeah.

[several agree]

Teacher [to the team]: OK? You've chosen that, OK. Fossil fuel abundant world! Energy prices collapse, we're swimming in a sea of oil and gas. Peak oil is all false and a bad rumour.

[The interface switches to adaptation scenario selection]

High adaptation potential

And if we go for high adaptation we can go 20% of our spend.

Yeah, definitely.

You want to try that?

Sounds good

[many agree]

Teacher: are you done?

Yeah

Teacher: OK so let's see

We should do a lot of mitigation as well

[unintelligible]

Teacher [to the class]: We have a completely different set of mitigation economics in this round, chosen by Betelgeuse, and an incredibly different set of adaptation mathematics as well. So let's play this final round and see who wins.

Our winning conditions?

Teacher [to the class]: Our winning condition is still the highest income.

Appendix B.

How to use CLOVER

This appendix aims to complement the theoretical explanations in Chapter 8, by showing a practical example.

As mentioned earlier, the goal of the CLOVER framework is to guide the design and evaluation of whole-class, collaborative activities. Moreover, the framework can potentially be used in various ways, and at various stages of a project. The example below illustrates one of the possible use cases of CLOVER, namely to guide the conversation during an initial design meeting. While the following story is fictional, it comprises a number of questions and concerns that have come up in actual studies and workshops I have conducted.

B.1. Letting CLOVER guide a new design

Suppose a teacher and a software developer are considering the possibility of designing a whole-classroom simulation together. Having seen a 1-minute demo video of 4Decades, both are feeling inspired and excited about the possibilities. However, none of them have any prior experience with designing technology for classroom use.

T: Have you read that guy's dissertation that I sent you? About classroom simulations?

D: Are you kidding? That thing is like 400 pages! I have a life, you know.

T: Yeah, same here. It's bizzarre. I skimmed through the whole thing - turns out the only truly useful bits are the last 5 pages or so, Appendix B. The idiot should have mentioned that in the abstract.

D: Pfff... what a null pointer. Alright then, lemme read that last bit please.

T: Sure, I'm gonna grab some tea while you're reading. Want some?

D: Yep, thanks.

(fast forward 5 minutes)

T: Here you go.

D: Cheers. Hey, shall we print that diagram after page 362?

T: Done.

D: You're a star. What is it anyway?

T: It's pretty much the entire thesis on one page, minus all the waffle about evidence and so forth.

D: Makes sense.

T: Totally.

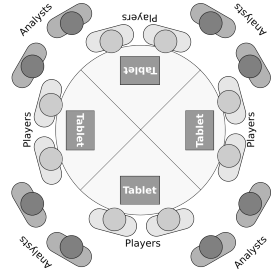
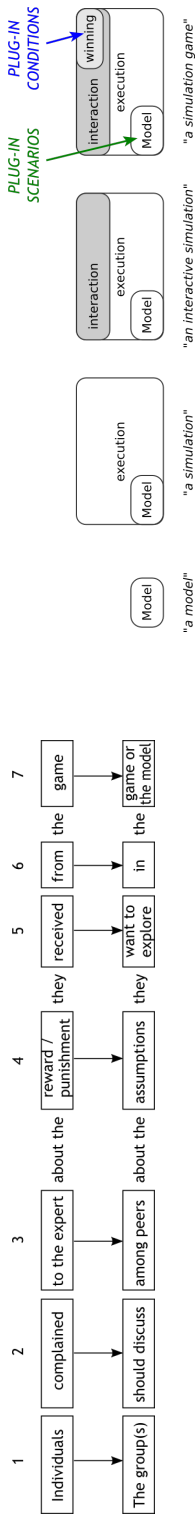
D: Wanna walk me through it?

T: Sure, you see the CLOVER tools are numbered 1-8.

CLOVER

Tools

Classroom Orchestration for Verbal Engagement and Reflection



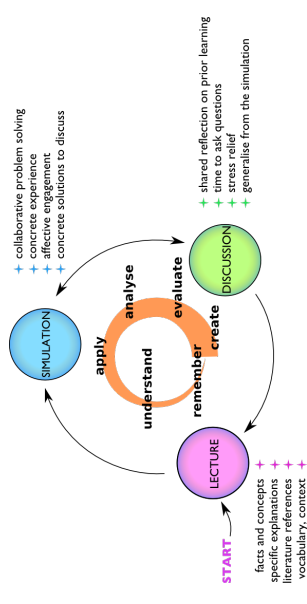
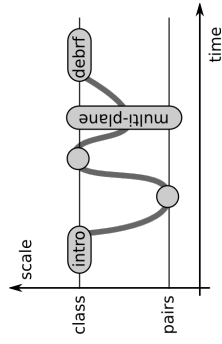
Benefits of Rule-Changing

Model Sim Game + Plugins

Lecture Sim Discussion

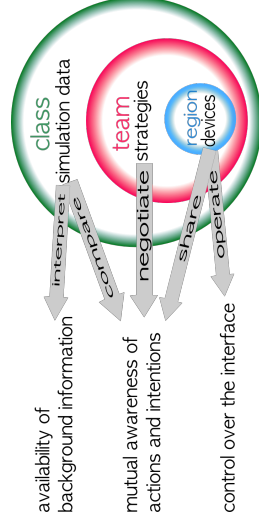
Spatial Formations

Multi-Plane Groupwork



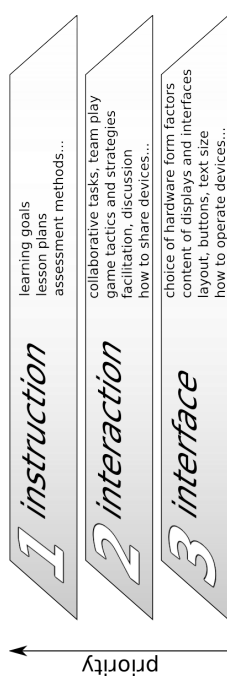
Triple-i Hierarchy

Three Bubbles



Gameplay Objectives

- 1 apply knowledge
- 2 instantly usable interfaces for 20-30
- 3 single teacher
- 4 reflect on assumptions via decisions
- 5 refresh by moving and talking
- 6 team incentives for negotiation
- 7 social mechanisms and play
- 8 ambient visualisations for awareness
- 9 inclusion and equitable accessibility



D: What do you mean tools? I thought we're designing a tool here? Now I'm confused.

T: Conceptual tools. To help us design.

D: Ah I get it. Sorry, go ahead.

T: Number one is basically just definitions, like what's the difference between a model, simulation...

D: ...and game, I get it. Plugin scenarios and winning conditions. Obvious. Nice diagram.

T: Really? This diagram took me a while to get my head around.

D: Haha, that's because you're not a developer!

T: That's *why* I'm not a developer, silly. Stop distracting, now let's get some work done here. My tea is growing a mustache.

D: So is mine. Weird.

T: Speaking of growing things - exponential growth seems to be really difficult to understand for a lot of kids I've been teaching.

D: Have you tried showing them a diagram?

T: Yes of course. And the formula. But it's all abstract magic to them. It would be fantastic if we could give them something tangible that they can play with and discuss.

D: Like a jar of bacteria? Those grow exponentially.

T: Cool idea. How about something less dangerous?

D: Like a virtual jar of bacteria? I guess I see where you're going with this.

T: Would that be difficult to model?

D: Depends. The numbers are just one line of code. But if you want 3D graphics...

T: Let's try to stick with numbers for the time being.

D: Why? I've got this really cool 3D engine for tablets that also does multitouch. It's amazing! You could make the bacteria wobble for example.

T: What are my students supposed to learn from that?

D: Well, I guess that bacteria can wobble. Or something. Right?

T: Let's try not to rush ahead. Look here, tool number 3 says you should first think about the learning goals before getting your head into interface details.

D: Yeah, I guess that makes sense. You're the teacher so I trust your expertise in that.

T: See, that's why we're doing this together.

D: Tell me then, what are the learning goals?

T: Well, I suppose if they had a discussion about what's happening then it would make them think harder and deeper about the topic.

D: What's there to discuss about bacteria? They grow until the glass is full and then they stop growing. Am I wrong?

T: Well, let's say the class could discuss about how long it will take to fill the glass.

D: ...or they could simply do the equation that will tell them the answer. Seems like a basic math test to me. If you turned that question into a game then you'd end up with the math-savvy kids always winning. And for the others the game would just suck, wouldn't it?

T: Fair point.

D: See, that's why we're doing this together.

T: :-P

D: Seriously though, I think we're stuck at this point. Exponential growth obviously doesn't work as a topic for discussion. Because the results are so clear and objective. It isn't really a matter of debate. I mean you can predict, you can estimate, but you can't really discuss. So I don't think this works as a topic.

T: Wait a minute. You're right, of course, about the results being predictable. One can't really debate how the formula works. But I suppose one can debate what its implications are. In a real-world context, think about the politics around overpopulation, economics and so on, that's where debates are happening right now.

D: So, potentially, you could link the game to some topical news article or video.

T: Maybe, why not. That would add some real pedagogical value in my opinion.

D: Nice. I'm going to add video streaming so they can watch the news on their iPads whilst they're playing the game.

T: Actually, I was thinking more along the lines of: First show them the news content and explain a few things about it. Then afterwards they play the game and discuss. Because look here, tool number 2 - the one we skipped earlier. It suggests, and I agree, to start with a bit of lecture (that's where the video belongs) and then to cycle mainly between simulation and discussion. Pedagogically, it just makes sense.

D: Why is that? What does the diagram say? The text is so tiny I can barely read it.

T: The short answer is trust me on this one. But of course, you can look up the detailed explanations in Chapter 8 and in the literature review if you're curious.

D: I'm not, actually, thanks for sparing me the details. Let's move on.

D: What's next on the list?

T: Tool number 4, "gameplay objectives".

D: This one makes perfect sense to me after seeing the 4Decades video.

T: Same here.

D: Just reading the list gives me an overall idea of what I'm aiming for, in terms of user interaction. It doesn't tell me how, but it tells me the goals.

T: Exactly. The how is in the recommendations.

D: The what?

T: Oh, I forgot to mention. There is a whole bunch of little design tips in Chapter 8, that is also part of CLOVER. It's just a few pages of compact information. You should have a look.

D: I will. Me likes compact information. What's tool number 5 - "three bubbles"?

T: That's another great one, but let's leave it for next time. I'm afraid I have to rush off soon. But I think we had a pretty good start in this short time.

D: I agree. It was fun. Let's continue this soon. In the meantime, I'm going to have a look at those recommendations you mentioned.

T: Great. And let's both keep an eye out for interesting reports or documentaries to do with exponential growth.

D: Doesn't have to be bacteria, right? Could be anything.

T: Yep.

D: Wow, this CLOVER thing has potential. It seems to have kept us focused.

T: It seems to have kept YOU focused, professor 3D multiwobble engine.

D: And you, doctor math quiz. Anyway, I'm sure we'll soon be rich and famous with this.

T: Do you think so? I would say we'd be lucky to make it into the appendix.

D: Yay! Go team appendix!

T: Sounds wrong.

D: You're right.

T: See you next time.

D: Thanks for the tea, my friend. Goodbye.

B.2. Other ways to use CLOVER

As mentioned earlier, CLOVER is not a monolithic framework, but rather a set of components that can be used more or less independently. I have used it in collaborative design with educators, including the two studies described in Chapter 9. One of them (UniPad) shows how one can design a new application from scratch, by sticking closely to the tools and recommendations. The other study (Comfy Birds) used CLOVER selectively for the purpose of extending an existing, single-screen application to the whole-classroom level. In addition to creating ideas, my collaborators and I have used CLOVER for reality checking during iterative design.

The modular nature of CLOVER implies that it can be modified by adding or removing components. Some components of the current version are specific to simulations and were therefore not used in the (non-simulation) ComfyBirds design. Conversely, one might add further tools and recommendations to the framework. This dissertation has touched on a wide range of concepts from the fields of education, modelling, HCI and game design, many of which could potentially be formulated as plugins to CLOVER.

As a result, the future might produce a variety of CLOVER versions. For instance, suppose the current version was renamed "Simulation CLOVER", then conceivably there could also be "Vocabulary CLOVER", "Creativity CLOVER", and so on.

Bibliography

- E. Ackermann. Perspective-taking and object construction: two keys to learning. *Constructionism in practice: designing, thinking, and learning in a digital world*, Lawrence Erlbaum, Mahwah, NJ, pages 25–35, 1996.
- I. AlAgha, A. Hatch, L. Ma, and L. Burd. Towards a teacher-centric approach for multi-touch surfaces in classrooms. In *ITS Tabletop*, pages 187–196. ACM, 2010.
- H. Alavi, P. Dillenbourg, and F. Kaplan. Distributed awareness for class orchestration. *Proceedings of the Third European Conference on Technology-Enhanced Learning (ECTEL09)*, 2009.
- V. Ambrosini and C. Bowman. Tacit knowledge: Some suggestions for operationalization. *Journal of Management Studies*, 38(6):811–829, 2001.
- L. Andersson, J. A. Olsson, B. Arheimer, and A. Jonsson. Use of participatory scenario modelling as platforms in stakeholder dialogues. *Water SA: Special HELP Edition: Special Issue 4*, 34:439–447, 2008.
- S. Arnab, R. Berta, J. Earp, S. de Freitas, M. Popescu, M. Romero, I. Stanescu, and M. Usart. Framing the Adoption of Serious Games in Formal Education. *Electronic Journal of e-Learning*, 10(2):159–171, 2012.
- E. Aronson, N. Blaney, C. Stephin, J. Sikes, and M. Snapp. *The jigsaw classroom*. Sage, 1978.

- K. Bachour, F. Kaplan, and P. Dillenbourg. An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies*, 3(3):203–213, 2010.
- O. Barreteau. Our companion modelling approach. *Journal of Artificial Societies and Social Simulation*, 6(2), 2003.
- S. Benford, D. Rowland, M. Flintham, R. Hull, J. Reid, J. Morrison, K. Facer, and B. Clayton. Savannah: Designing a location-based game simulating lion behaviour. In *International Conference on Advances in Computer Entertainment Technology*, 2004.
- D. Bennet and A. Bennet. Engaging tacit knowledge in support of organizational learning. *VINE*, 38(1):72–94, 2008.
- A. Blandford and D. Furniss. DiCoT: a methodology for applying distributed cognition to the design of team working systems. *Cognition*, 3941:26–38, 2006.
- P. Blatchford, P. Bassett, and P. Brown. Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, 21(6):715–730, Dec. 2011.
- B. S. Bloom and D. R. Krathwohl. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. 1956.
- H. Brignull and Y. Rogers. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT*, pages 17–24, 2003.
- J. P. Briot, P. Guyot, and M. Irving. Participatory simulation for collective management of protected areas for biodiversity conservation and social inclusion. *AI, Simulation and Planning in High Autonomy Systems (AIS) & Conceptual Modeling and Simulation (CMS)*, *International Modeling and Simulation Multiconference*, pages 183–188, 2007.

- J. S. Brown, A. Collins, and P. Duguid. Situated cognition and the culture of learning. *Educational researcher*, 18(1):32, 1989.
- J. S. Bruner. *The process of education*. Harvard University Press, 2009.
- J. C. Castella, T. N. Trung, and S. Boissau. Participatory simulation of land-use changes in the northern mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. *Ecology and Society*, 10(1):27, 2005.
- W. Clark and R. Luckin. ipads in the classroom. *What The Research Says*, 2013.
- R. Cober, C. Fong, A. Gnoli, B. L. Silva, M. Lui, C. Madeira, C. McCann, T. Moher, J. Slotta, and M. Tissenbaum. Embedded Phenomena for Knowledge Communities: Supporting complex practices and interactions within a community of inquiry in the elementary science classroom. In *Proceedings of the Tenth International Conference of the Learning Sciences*, volume 2, pages 64–71, 2012.
- V. Colella. Participatory Simulations: Building Collaborative Understanding through Immersive Dynamic Modeling. *The Journal of the Learning Sciences*, 9(4):pp. 471—500, 2000.
- S. Cox, N. Pattinson, and T. Woodward. Principles of learning, 2014. URL <http://curve.coventry.ac.uk/cu/file/33362904-d31d-ad7a-a7ff-b5f63bca5bc2/1/studyskills.zip/ehe/StudWork/Prinlers/prinlers.html>.
- C. K. W. De Dreu and M. a. West. Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology*, 86(6):1191–1201, 2001.
- S. De Freitas and M. Oliver. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3):249–264, 2006.

- M. S. A. De Hei, J.-W. Strijbos, E. Sjoer, and W. Admiraal. Collaborative learning in higher education: lecturers' practices and beliefs. *Research Papers in Education*, (May):1–16, Apr. 2014.
- T. De Jong and W. R. Van Joolingen. Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research*, 68(2):179, 1998.
- J. Dewey. *How we think*. Heath, 1933.
- P. Dietz and D. Leigh. Diamondtouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226. ACM, 2001.
- P. Dillenbourg. What do you mean by collaborative learning. *Collaborative learning: Cognitive and computational approaches*, pages 1–16, 1999.
- P. Dillenbourg. Over-scripting cscl: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL?*, pages 61–91, 2002.
- P. Dillenbourg and P. Jermann. Technology for classroom orchestration. *New science of learning*, pages 1–20, 2010.
- P. Dillenbourg, M. Baker, A. Blaye, and C. O'Malley. The evolution of research on collaborative learning. *Learning in Humans and Machine: Towards an interdisciplinary learning science*, pages 189–211, 1996.
- P. Dillenbourg, G. Zufferey, H. Alavi, P. Jermann, S. Do-Lenh, Q. Bonnard, S. Cuendet, and F. Kaplan. Classroom orchestration: The third circle of usability. *CSCL2011 Proceedings*, 1:510–517, 2011.
- P. Dillenbourg, M. Nussbaum, Y. Dimitriadis, and J. Roschelle. Design for classroom orchestration. *Computers & Education*, 2012.

- A. A. DiSessa. *On "Learnable" Representations of Knowledge: A Meaning for the Computational Metaphor*. Massachusetts Institute of Technology, A. I. Laboratory, 1977.
- W. Doise, G. Mugny, and A. Saint James-Emler. *The social development of the intellect*. Pergamon Press Oxford, 1984.
- M. Dunleavy, C. Dede, and R. Mitchell. Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18(1):7–22, 2009.
- A. Echeverría, C. García-Campo, M. Nussbaum, F. Gil, M. Villalta, M. Améstica, and S. Echeverría. A framework for the design and integration of collaborative classroom games. *Computers & Education*, 57(1):1127–1136, Aug. 2011.
- S. Egenfeldt-Nielsen. Practical barriers in using educational computer games. *On the Horizon*, 12(1):18–21, 2004.
- S. Egenfeldt-Nielsen. Overview of research on the educational use of video games. *Digital Kompetanse*, 1(3):184–213, 2006.
- J. B. Ellis, K. Luther, K. Bessiere, and W. A. Kellogg. Games for virtual team building. In *Proceedings of the 7th ACM conference on Designing interactive systems*, pages 295–304. ACM, 2008.
- Y. Engeström. Learning by expanding: An activity-theoretical approach to developmental research. 1987.
- P. A. Enkvist, J. Dinkel, and C. Lin. Impact of the Financial Crisis on Carbon Economics: Version 2.1 of the Global Greenhouse Gas Abatement Cost Curve. *McKinsey & Company*, 2010.

- P. Eykhoff. *System identification : parameter and state estimation*. Wiley-Interscience, London {;;New} York [etc.], 1974.
- R. M. Fanning and D. M. Gaba. The role of debriefing in simulation-based learning. *Simulation in Healthcare*, 2(2):115, 2007.
- A. J. Faria and W. J. Wellington. A survey of simulation game users, former-users, and never-users. *Simulation & Gaming*, 35(2):178, 2004.
- A. J. Faria, D. Hutchinson, W. J. Wellington, and S. Gold. Developments in Business Gaming. *Simul. Gaming*, 40(4):464–487, Aug. 2009.
- M. R. Fellenz. Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment and Evaluation in Higher Education*, 29(6):703–719, 2004.
- M. Flanagan. *Critical play: radical game design*. MIT press, 2009.
- R. Gacitua, L. Ma, B. Nuseibeh, P. Piwek, A. de Roeck, M. Rouncefield, P. Sawyer, A. Willis, and H. Yang. Making tacit requirements explicit. In *Managing Requirements Knowledge (MARK), 2009 Second International Workshop on*, pages 40–44, Sept 2009.
- I. A. Games and K. Squire. Design thinking in gamestar mechanic: the role of gamer experience on the appropriation of the discourse practices of game designers. In *Proceedings of the 8th international conference on International conference for the learning sciences - Volume 1*, ICLS'08, pages 257–264. International Society of the Learning Sciences, 2008.
- H. Gardner. *The unschooled mind: How children think and how schools should teach*. Basic Books, 1991.
- J. P. Gee. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1):20, 2003.

- J. P. Gee. Are video games good for learning? *Worlds in Play: International Perspectives on Digital Games Research*, 21:323, 2007.
- J. P. Gee. Game-like learning: An example of situated learning and implications for opportunity to learn. *Assessment, equity, and opportunity to learn*, pages 200–221, 2008.
- A. Gnoli, A. Perritano, P. Guerra, B. Lopez, J. Brown, and T. Moher. Back to the future: embodied classroom simulations of animal foraging. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, pages 275–282. ACM, 2014.
- S. Greenberg, N. Marquardt, T. Ballendat, R. Diaz-Marino, and M. Wang. Proxemic interactions: the new ubicomp? *interactions*, 18(1):42–50, Jan. 2011. ISSN 1072-5520. doi: 10.1145/1897239.1897250. URL <http://doi.acm.org/10.1145/1897239.1897250>.
- E. T. Hall. *Proxemics: The study of man's spatial relations*. 1962.
- G. Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.
- A. Harris, J. Rick, V. Bonnett, N. Yuill, R. Fleck, P. Marshall, and Y. Rogers. Around the Table : Are Multiple-Touch Surfaces Better Than Single-Touch for Children's Collaborative Interactions? In *Proceedings of the 9th international conference on Computer supported collaborative learning, CSCL*, pages 335–344, 2009.
- E. R. Hayes and J. P. Gee. No Selling the Genie Lamp: a game literacy practice in The Sims. *E-Learning and Digital Media*, 7(1):67–78, 2010.
- E. R. Hayes, J. P. Gee, I. Games, R. J. Torres, K. Peppler, Y. B. Kafai, N. Pinkard, E. Klopfer, H. Scheintaub, M. Rogers, K. S. Forssell, C. K. Martin, B. Barron, W. Eugene, S. B. Daily, U. Acholonu, L. Takeuchi, S. Walter, and K. Briggs. New perspectives on learning through (game) design. In *Proceedings of the 8th international*

- conference on International Conference for the Learning Sciences - Volume 3, ICLS'08*, pages 253–257. International Society of the Learning Sciences, 2008.
- S. Higgins, E. Hall, K. Wall, P. Woolner, and C. McCaughey. The impact of school environments: A literature review. *The Centre for Learning and Teaching, School of Education, Communication and Language Science, University of Newcastle*. Accessed online on, 10:04–08, 2005.
- S. Higgins, E. Mercier, L. Burd, and A. Joyce-gibbons. Multi-touch tables and collaborative learning _1259 1041..1054. 43(6):1041–1054, 2012.
- S. E. Higgins, E. Mercier, E. Burd, and A. Hatch. Multi-touch tables and the relationship with collaborative classroom pedagogies: A synthetic review. *International Journal of Computer-Supported Collaborative Learning*, 6(4):515–538, Sept. 2011.
- M. L. Hilton and M. A. Honey. *Learning Science Through Computer Games and Simulations*. The National Academies Press, 2011.
- P. Hollins and N. Whitton. From the Games Industry: Ten Lessons for Game-Based Learning. *International Journal of Virtual and Personal Learning Environments (IJVPLE)*, 2(2):73–82, 2011.
- C. Hoyles, R. Noss, and R. Adamson. Rethinking the microworld idea. *Journal of educational computing research*, 27(1):29–53, 2002.
- E. Hutchins. The social organization of distributed cognition. *Perspectives on Socially Shared Cognition*, 1991.
- C. Inman, V. H. Wright, and J. A. Hartman. Use of Second Life in K-12 and higher education: A review of research. *SECOND LIFE AS A LANGUAGE LEARNING TOOL (EFL)*, page 67, 2010.

- H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. *Proceedings of the ACM SIGCHI Conference on ...*, (March):234–241, 1997.
- B. Z. Johnson. *Virtual economies as financial literacy sandboxes: case study in Gaia Online*. PhD thesis, University of Minnesota, 2012.
- D. Jonassen and J. Strobel. Modeling for meaningful learning. *Engaged learning with emerging technologies*, pages 1–27, 2006.
- D. H. Jonassen. Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational technology research and development*, 39(3):5–14, 1991.
- D. H. Jonassen and P. Henning. Mental models: Knowledge in the head and knowledge in the world. In *Proceedings of the 1996 international conference on Learning sciences*, pages 433–438. International Society of the Learning Sciences, 1996.
- D. H. Jonassen and B. Kim. Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology Research and Development*, 58(4):439–457, 2010.
- D. A. Jones and M. Chang. Multiplayer Online Role Playing Game for Teaching Youth Finance in Canada. *International Journal of Online Pedagogy and Course Design (IJOPCD)*, 2(2):44–59, 2012.
- J. Juul. The open and the closed: Games of emergence and games of progression. In *CGDC Conf.*, 2002.
- J. Juul. The game, the player, the world: Looking for a heart of gameness. In *Level up: Digital games research conference proceedings*, volume 120, page 121. Utrecht Univ., Utrecht, Holland, 2003.

- Y. B. Kafai, W. Q. Burke, and D. A. Fields. What Videogame Making Can Teach Us About Access and Ethics in Participatory Culture. 2009.
- P. Kaminsky and D. Simchi-Levi. A new computerized beer game: A tool for teaching the value of integrated supply chain management. *Global Supply Chain and Technology Management, POMS Series in Technology and Operations Management*, 1:216–225, 1998.
- A. B. Kayes, D. C. Kayes, and D. A. Kolb. Experiential learning in teams. *Simulation & Gaming*, 36(3):330–354, 2005.
- C. J. Keller, N. D. Finkelstein, K. K. Perkins, and S. J. Pollock. Assessing the Effectiveness of a Computer Simulation in Introductory Undergraduate Environments. In *AIP Conference Proceedings*, pages 121–124. AIP, Jan. 2007.
- H. Kelly, K. Howell, E. Glinert, L. Holding, C. Swain, A. Burrowbridge, and M. Roper. How to build serious games. *Commun. ACM*, 50(7):44–49, July 2007.
- A. Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.
- A. Kharrufa, M. Balaam, P. Heslop, D. Leat, P. Dolan, and P. Olivier. Tables in the Wild : Lessons Learned from a Large-Scale Multi-Tabletop Deployment. In *Proc. of CHI '13*, Paris, France, 2013. ACM.
- P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.
- E. Klopfer and K. Squire. Environmental detectives—the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development*, 56(2):203–228, 2008.

- E. Klopfer, K. Squire, and H. Jenkins. Environmental Detectives: PDAs as a window into a virtual simulated world. In *Wireless and Mobile Technologies in Education, 2002. Proceedings. IEEE International Workshop on*, pages 95–98. IEEE, 2002.
- E. Klopfer, S. Yoon, and J. Perry. Using palm technology in participatory simulations of complex systems: A new take on ubiquitous and accessible mobile computing. *Journal of Science Education and Technology*, 14(3):285–297, 2005.
- E. Klopfer, H. Scheintaub, W. Huang, and D. Wendel. StarLogo TNG: Making Agent-Based Modeling Accessible and Appealing to Novices. *Artificial Life Models in Software*, page 151, 2009.
- D. A. Kolb and Others. *Experiential learning: Experience as the source of learning and development*. Prentice-Hall Englewood Cliffs, NJ, 1984.
- M. Krain and J. S. Lantis. Building knowledge? Evaluating the effectiveness of the global problems summit simulation. *International Studies Perspectives*, 7(4):395–407, 2006.
- D. R. Krathwohl. A revision of bloom’s taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.
- S. Kreitmayer, S. Peake, R. Laney, and Y. Rogers. Designing a Large Multi-Player Simulation Game to Encourage Reflection and Critical Debate. In *Proceedings of the 5th European Conference on Games Based Learning*, 2011.
- S. Kreitmayer, Y. Rogers, R. Laney, and S. Peake. From Participatory to Contributory Simulations: Changing the Game in the Classroom. In *Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI ’12*, ACM, 2012.
- S. Kreitmayer, R. Laney, S. Peake, and Y. Rogers. Sharing Bubbles: Reflections on Offline Multi-Surface Scenarios. In *Contribution to the SOFTec Workshop in Proceedings of UBIComp ’13*, page 4, Zurich, Switzerland, 2013a. ACM.

- S. Kreitmayer, Y. Rogers, R. Laney, and S. Peake. UniPad: Orchestrating Collaborative Activities Through Shared Tablets and an Integrated Wall Display. In *Proc. of UBIComp '13*. ACM, 2013b.
- K. Kuutti. Activity theory as a potential framework for human-computer interaction research. *Context and consciousness: Activity theory and human-computer interaction*, pages 17–44, 1996.
- T. Lainema. Perspective Making. *Simulation & Gaming*, 40(1):48, 2009.
- J. Lave. *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press, 1988.
- J. Lave and E. Wenger. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, 1991.
- M. Lipman. *Thinking in education*. Cambridge University Press, 2003.
- C.-K. Looi and Y. Toh. Orchestrating the Flexible Mobile Learning Classroom. *Increasing Access through Mobile Learning*, pages 161–174, 2014.
- C.-K. Looi, W. Chen, and F.-K. Ng. Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness. *Computers & Education*, 54(1): 14–26, 2010.
- M. Lui, A. Kuhn, A. Acosta, C. Quintana, and J. D. Slotta. Supporting learners in collecting and exploring data from immersive simulations in collective inquiry. *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2103–2112, 2014.
- L. Lyons, J. Lee, C. Quintana, and E. Soloway. Preliminary evaluation of a synchronous co-located educational simulation framework. In *Proc. of CHI '06*, CHI EA '06, pages 1055–1060, New York, NY, USA, 2006. ACM.

- P. Marshall, R. Morris, Y. Rogers, S. Kreitmayer, and M. Davies. Rethinking 'Multi-user': an In-the-Wild Study of How Groups Approach a Walk-Up-and-Use Tabletop Interface. 2011.
- R. Martinez-Maldonado, J. Kay, and K. Yacef. Orchestrating a Multi-tabletop Classroom : From Activity Design to Enactment and Reflection. In *Proc. of the 2012 ACM intl. conf. on Interactive tabletops and surfaces ITS*, pages 119–128, 2012.
- R. Martinez-Maldonado, Y. Dimitriadis, A. Clayphan, J. a. Muñoz Cristóbal, L. P. Prieto, M. J. Rodríguez-Triana, and J. Kay. Integrating orchestration of ubiquitous and pervasive learning environments. *Proceedings of the 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration - OzCHI '13*, pages 189–192, 2013.
- R. E. Mayer. Rote versus meaningful learning. *Theory into practice*, 41(4):226–232, 2002.
- J. McGonigal. *Reality is broken: Why games make us better and how they can change the world*. Penguin, 2011.
- D. L. Meadows. Learning to be simple: My odyssey with games. *Simulation & Gaming*, 30(3):342–351, 1999.
- D. L. Meadows, D. E. Shannon, T. Fiddaman, and U. of New Hampshire. *Fish Banks: A Microcomputer Assisted Simulation that Teaches Principles for Sustainable Management of Renewable Resources*. Institute for policy and social science research, 1989.
- N. Mercer. The Social Brain, Language, and Goal-Directed Collective Thinking: A Social Conception of Cognition and Its Implications for Understanding How We Think, Teach, and Learn. *Educational Psychologist*, 48(3):148–168, July 2013.
- E. M. Mercier, S. E. Higgins, and A. Joyce-Gibbons. The effects of room design on computer-supported collaborative learning in a multi-touch classroom. *Interactive Learning Environments*, (ahead-of-print):1–19, 2014.

- T. Moher. Embedded phenomena: supporting science learning with classroom-sized distributed simulations. In *Proc. of CHI '06*, pages 691–700. ACM, 2006. ISBN 1595933727.
- T. Moher. Learning and participation in a persistent whole-classroom seismology simulation. In *Proceedings of the 8th international conference on International conference for the learning sciences - Volume 2*, ICLS'08, pages 82–90. International Society of the Learning Sciences, 2008.
- T. Moher, S. Hussain, T. Halter, and D. Kilb. RoomQuake: embedding dynamic phenomena within the physical space of an elementary school classroom. In *CHI '05 extended abstracts*, pages 1665–1668. ACM, 2005.
- T. Moher, B. Uphoff, D. Bhatt, B. López Silva, and P. Malcolm. WallCology. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 163, New York, New York, USA, Apr. 2008. ACM Press. ISBN 9781605580111. doi: 10.1145/1357054.1357082.
- T. Moher, J. Wiley, A. Jaeger, B. L. Silva, and D. Kilb. Spatial and temporal embedding for science inquiry: an empirical study of student learning. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1*, pages 826–833. International Society of the Learning Sciences, 2010.
- A. Monroy-Hernández and M. Resnick. FEATURE: Empowering kids to create and share programmable media. *interactions*, 15(2):50–53, Mar. 2008.
- J. Moon. *Short Courses & Workshops: Improving the Impact of Learning, Training & Professional Development*. Psychology Press, 2001.
- A. L. Morgan. Toward a global theory of mind: The potential benefits of presenting a range of IR theories through active learning. *International Studies Perspectives*, 4(4): 351–370, 2003.

- M. R. Morris. Designing tabletop groupware. In *UIST'05 Doctoral Symposium*, 2005.
- M. R. Morris, J. Lombardo, and D. Wigdor. WeSearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 401–410. ACM, 2010.
- G. Moss, C. Jewitt, R. Levañiç, V. Armstrong, A. Cardini, F. Castle, B. Allen, A. Jenkins, and M. Hancock. The Interactive Whiteboards , Pedagogy and Pupil Performance Evaluation : An Evaluation of the Schools Whiteboard Expansion (SWE) Project : London Challenge The Interactive Whiteboards , Pedagogy. 2007.
- F. Novellis and T. Moher. How real is real enough?: designing artifacts and procedures for embodied simulations of science practices. pages 90–98, 2011.
- E. Nussbaum. Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, 33(3):345–359, July 2008.
- H. F. O’Neil, R. Wainess, and E. L. Baker. Classification of learning outcomes: Evidence from the computer games literature. *Curriculum journal*, 16(4):455–474, 2005.
- B. O’Neill. Financial simulations for young adults: Making the "real world" real. *Journal of Extension*, 46(6), 2008.
- M. Overmars. Teaching computer science through game design. *Computer*, 37(4):81–83, 2004.
- S. Papert. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc., 1980.
- S. Papert and I. Harel. Situating constructionism. *Constructionism*, 1991a.
- S. Papert and I. Harel. Situating constructionism. *Constructionism*, 36:1–11, 1991b.
- J. Piaget and B. Inhelder. *The Psychology Of The Child*. Basic Books, 1969.

- A. M. Piper, E. O'Brien, M. R. Morris, and T. Winograd. SIDES: a cooperative tabletop computer game for social skills development. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 1–10. ACM, 2006.
- L. Prieto, M. Holenko Dlab, I. Gutiérrez, M. Abdulwahed, and W. Balid. Orchestrating technology enhanced learning: a literature review and a conceptual framework. . . . *Technology Enhanced* . . . , 3(6), 2011.
- L. Prieto, S. Villagr a-Sobrino, Y. Dimitriadis, and I. Jorr n-Abell n. Orchestrating classroom cscl: a multi-level pattern approach for design and enactment. *Handbook of Design in Educational Technology*, eds R. Luckin, et al., Routledge, New York, pp. 112 123, 2013.
- L. P. Prieto. *Supporting Orchestration of Blended CSCL Scenarios in Distributed Learning Environments*. PhD thesis, 2012.
- A. Repenning and A. Ioannidou. Broadening participation through scalable game design. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education*, pages 305–309. ACM, 2008.
- M. Resnick. Distributed constructionism. In *Proceedings of the 1996 international conference on Learning sciences*, pages 280–284. International Society of the Learning Sciences, 1996.
- M. Resnick. *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*. The MIT Press, 1997.
- M. Resnick and U. Wilensky. Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *Journal of the Learning Sciences*, 7(2): 153–172, 1998.

- J. W. Rice. New media resistance: Barriers to implementation of computer video games in the classroom. *Learning and Teaching with Electronic Games*, page 249, 2009.
- J. Rick. Tiny tabletops: A research agenda. In *Workshop Proceedings of ICLS 2012 "Digital Ecosystems for Collaborative Learning"*, 2012a.
- J. Rick. Tiny tabletops: A research agenda. In *Workshop Proceedings of ICLS 2012 "Digital Ecosystems for Collaborative Learning"*, 2012b.
- J. Rick and P. Marshall. Towards a Constructivist Pedagogy in the Ubicomp Classroom. pages 4–7, 2010.
- L. P. Rieber. Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development*, 44(2):43–58, 1996.
- H. W. J. Rittel and M. M. Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973.
- U. Ritterfeld, M. Cody, and P. Vorderer. *Serious games: Mechanisms and effects*. Routledge, 2010.
- Y. Rogers. Hci theory: classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, 5(2):1–129, 2012.
- Y. Rogers, N. Yuill, and P. Marshall. Contrasting Lab-based and In-the-wild Studies for Evaluating Multi-user Technologies. . . . of *Digital Technology Research*, pages 359–373, 2013.
- J. Roschelle and R. Pea. A walk on the WILD side: How wireless handhelds may change CSCL. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, pages 51–60. International Society of the Learning Sciences, 2002.

- J. Roschelle, Y. Dimitriadis, and U. Hoppe. Classroom orchestration: synthesis. *Computers & Education*, 69:523–526, 2013.
- J. M. Roschelle, R. D. Pea, C. M. Hoadley, D. N. Gordin, and B. M. Means. Changing How and What Children Learn in School with Computer-Based Technologies. *The Future of Children*, 10(2):pp. 76–101, 2000. ISSN 10548289.
- K. Salen. Gaming literacies: A game design study in action. *Journal of Educational Multimedia and Hypermedia*, 16(3):301–322, 2007.
- K. Salen and E. Zimmerman. *Rules of play: Game design fundamentals*. The MIT Press, 2004.
- W. A. Sandoval and B. J. Reiser. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3):345–372, 2004.
- N. B. Sardone. Exploring Personal Finance Concepts Through Simulation Role Play. pages 1–18, 2010.
- G. L. Savoldelli, V. N. Naik, J. Park, H. S. Joo, R. Chow, and S. J. Hamstra. Value of debriefing during simulated crisis management: oral versus video-assisted oral feedback. *Anesthesiology*, 105(2):279–285, 2006.
- S. Scott and M. Sheelagh. Territoriality in collaborative tabletop workspaces. *Proceedings of the 2004 . . .*, pages 294–303, 2004.
- M. Sharples. Shared orchestration within and beyond the classroom. *Computers and Education*, 69:504–506, 2013.
- S. M. Shellman and K. Turan. Do simulations enhance student learning? An empirical evaluation of an IR simulation. *Journal of Political Science Education*, 2(1):19–32, 2006.

- N. Shin, D. H. Jonassen, and S. McGee. Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of research in science teaching*, 40(1):6–33, 2003.
- C. Silvia. The Impact of Simulations on Higher-Level Learning. *Journal of Public Affairs Education*, pages 397–422, 2012.
- J. Slotta. Evolving the classrooms of the future: The interplay of pedagogy, technology and community. *Classroom of the future: Orchestrating collaborative spaces*, pages 215–242, 2010.
- O. Smørðal, J. Slotta, T. Moher, F. Novellis, A. Gnoli, B. Lopez Silva, M. Lui, A. Jor-net, C. Jahreie, and I. Krange. Hybrid spaces for science learning: New demands and opportunities for research. In *International Conference of the Learning Sciences. Sydney, Australia*. http://www.uv.uio.no/intermedia/english/research/projects/miracle/news/pdf/ICLS_2012_Symposium_Hybrid_Spaces.pdf, 2012.
- K. Squire and S. A. Barab. Replaying history: Learning world history through playing Civilization III. *Indiana University, Indianapolis, IN*, 2004.
- K. D. Squire et al. *Replaying history: Learning world history through playing Civilization III*. Indiana University Bloomington, IN, 2004.
- G. Stahl. *Group cognition: Computer support for building collaborative knowledge*. MIT Press Cambridge, MA, 2006.
- G. Stahl, T. Koschmann, and D. Suthers. Computer-supported collaborative learning: An historical perspective. *Cambridge handbook of the learning sciences*, 2006, 2006.
- C. Steinkuehler and B. Z. Johnson. Computational literacy in online games: The social life of mods. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 1(1):53–65, 2009.

- J. D. Sterman, T. Fiddaman, T. Franck, A. Jones, S. McCauley, P. Rice, E. Sawin, and L. Siegel. Management flight simulators to support climate negotiations. *Environmental Modelling & Software*, 44(0):122–135, 2013.
- N. Stern. The economics of climate change. *World Economics*, 7(2):1–10, 2006.
- L. A. Suchman. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, 1987.
- I. Suto. 21st Century skills: Ancient, ubiquitous, enigmatic? *Research Matters: A Cambridge Assessment Publication*, 2013.
- M. Tomasello and Others. *Why we cooperate*, volume 206. MIT press Cambridge, MA, 2009.
- S. Turkle. From powerful ideas to PowerPoint. *Convergence: The International Journal of Research into New Media Technologies*, 9(2):19, 2003.
- M. B. Twidale. Over the Shoulder Learning: Supporting Brief Informal Learning. *Computer Supported Cooperative Work (CSCW)*, 14(6):505–547, Nov. 2005.
- I. Utting, S. Cooper, M. Kolling, J. Maloney, and M. Resnick. Alice, Greenfoot, and Scratch—A Discussion. *ACM Transactions on Computing Education (TOCE)*, 10(4):17, 2010.
- L. S. Vygotsky. Mind and society: The development of higher mental processes, 1978.
- P. Watzlawick, J. H. Beavin, and D. D. A. Jackson. *Pragmatics of human communication: A study of international patterns, pathologies, and paradoxes*. Norton, 1967.
- R. Wegerif. A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning*, 1(1):143–157, Mar. 2006.
- M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

- N. Whitton. *Learning with digital games*. Routledge, 2009.
- D. Wigdor and R. Balakrishnan. Empirical investigation into the effect of orientation on text readability in tabletop displays. In *Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work*, ECSCW'05, pages 205–224, New York, NY, USA, 2005. Springer-Verlag New York, Inc.
- U. Wilensky and K. Reisman. Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction*, 24(2):171–209, 2006.
- U. Wilensky and W. Stroup. Learning through participatory simulations: network-based design for systems learning in classrooms. In *Proceedings of the 1999 conference on Computer support for collaborative learning*, CSCL '99. International Society of the Learning Sciences, 1999.
- U. Wilensky and W. Stroup. Networked gridlock: Students enacting complex dynamic phenomena with the HubNet architecture. In *Fourth International Conference of the Learning Sciences*, pages 282–289. Citeseer, 2000.
- N. Yuill. Visiting joke city: How can talking about jokes foster metalinguistic awareness in poor comprehenders. *Reading comprehension strategies: Theories, interventions and technologies*, pages 325–346, 2007.
- N. Yuill. The relation between ambiguity understanding and metalinguistic discussion of joking riddles in good and poor comprehenders: Potential for intervention and possible processes of change. *First language*, 29(1):65–79, 2009.
- N. Yuill and J. Bradwell. The laughing pc: How a software riddle package can help children's reading comprehension. In *Proceedings of the British Psychological Society*, volume 6, page 119, 1998.

- N. Yuill and Y. Rogers. Mechanisms for Collaboration: A Design and Evaluation Framework for Multi-User Interfaces. *Transactions on Computer-Human Interaction, ACM*, 2012.
- N. Yuill, Y. Rogers, and J. Rick. Pass the iPad: collaborative creating and sharing in family groups. *Proceedings of the 2013 ACM annual ...*, pages 941–950, 2013.
- N. Yuill, S. Holt, C. Girvan, and S. Kreitmayer. Being Aware: Using Linked Tablets to Promote Collaboration and Communication between Children with ASC. In *Proceedings of the 2nd international conference on Innovative Technologies (IT) for Autism (ASD), ITASD*, 2014.